Psychiatry in pictures:

Psychiatry in pictures
ALLAN BEVERIDGE

Highlights of this issue:

Highlights of this issue
SUKHWINDER S. SHERGILL

EDITORIALS:

Does compulsory or supervised community treatment reduce ‘revolving door’ care?:
Legislation is inconsistent with recent evidence
STEPHEN KISELY and LESLIE Anne CAMPBELL

The medical model is dead – long live the medical model
PREMAL SHAH and DEBORAH MOUNTAIN

REVIEW ARTICLES:

Effects of treating postnatal depression on mother–infant interaction and child
development: Systematic review
AMUDHA S. POOBALAN, LORNA S. AUCOTT, LOUISE ROSS, W. CAIRNS. S. SMITH,
PETER J. HELMS, and JUSTIN H. G. WILLIAMS

Cortisol and post-traumatic stress disorder in adults: Systematic review and meta-analysis
MARIE-LOUISE MEEWISSE, JOHANNES B. REITSMA, GIEL-JAN DE VRIES,
BERTHOLD P. R. GERSONS, and MIRANDA OLFF
PAPERS:

Prevalence of depressive symptoms and syndromes in later life in ten European countries: The SHARE study
ERICO CASTRO-COSTA, MICHAEL DEWEY, ROBERT STEWART, SUBE BANERJEE, FELICIA HUPPERT, CARLOS MENDONCA-LIMA, CHRISTOPHE BULA, FRIEDEL REISCHE, JOHANNES WANCATA, KAREN RITCHIE, MAGDA TSOLAKI, RAIMUNDO MATEOS, and MARTIN PRINCE

Genotype effects of CHRNA7, CNR1 and COMT in schizophrenia: interactions with tobacco and cannabis use
STANLEY ZAMMIT, GILLIAN SPURLOCK, HYWEL WILLIAMS, NADINE NORTON, NIGEL WILLIAMS, MICHAEL C. O’DONOVAN, and MICHAEL J. OWEN

Psychophysiological and behavioural characteristics of individuals comorbid for antisocial personality disorder and schizophrenia-spectrum personality disorder
ROBERT A. SCHUG, ADRIAN RAINE, and RAND R. WILCOX

Impact and clinical significance of a preventive intervention for disruptive boys: 15-year follow-up
RACHEL BOISJOLI, FRANK VITARO, ÉRIC LACOURSE, EDWARD D. BARKER, and RICHARD E. TREMBLAY

Structured patient–clinician communication and 1-year outcome in community mental healthcare: Cluster randomised controlled trial
STEFAN PRIEBE, ROSEMARIE McCABE, JENS BULLENKAMP, LARS HANSSON, CHRISTOPH LAUBER, RAFAEL MARTINEZ-LEAL, WULF RÖSSLER, HANS SALIZE, BENGT SVENSSON, FRANCISCO TORRES-GONZALES, ROB VAN DEN BRINK, DURK WIERSMA, and DONNA J. WRIGHT

Clinical effectiveness of treatments for anorexia nervosa in adolescents: Randomised controlled trial
SIMON G. GOWERS, ANDREW CLARK, CHRIS ROBERTS, ALISON GRIFFITHS, VANESSA EDWARDS, CLAUDINE BRYAN, NICOLA SMETHEURST, SARAH BYFORD, and BARBARA BARRETT

Economic evaluation of a randomised controlled trial for anorexia nervosa in adolescents
SARAH BYFORD, BARBARA BARRETT, CHRIS ROBERTS, ANDREW CLARK, VANESSA EDWARDS, NICOLA SMETHEURST, and SIMON G. GOWERS
**Antidepressant effects of augmentative transcranial magnetic stimulation: Randomised multicentre trial**
UWE HERWIG, ANDREAS J. FALLGATTER, JACQUELINE HÖPPNER, GERHARD W. ESCHWEILER, MARTINA KRON, GÖRAN HAJAK, FRANK PADBERG, ANGELA NADERI-HEIDEN, BIRGIT ABLER, PETER EICHHAMMER, NICOLA GROSSHEINRICH, BIRGIT HAY, THOMAS KAMMER, BERTHOLD LANGGUTH, CHRISTOPH LASKE, CHRISTIAN PLEWNIA, MELANY M. RICHTER, MERTEN SCHULZ, STEFAN UNTERECKER, ANTONIA ZINKE, MANFRED SPITZER, and CARLOS SCHÖNFELDT-LECUONA

**SHORT REPORTS:**

**Intergenerational transmission of health beliefs in somatoform disorders: Exploratory study**
TAMSIN MARSHALL, DAVID P.H. JONES, PAUL G. RAMCHANDANI, ALAN STEIN, and CHRISTOPHER BASS

**Effects of training community staff in interventions for substance misuse in dual diagnosis patients with psychosis (COMO study): Cluster randomised trial**
SONIA JOHNSON, GRAHAM THORNICROFT, SARAH AFUAWE, MORVEN LESEE, IAN R. WHITE, ELIZABETH HUGHES, SHAMIL WANIGARATNE, HELEN MILES, and TOM CRAIG

**Are people with schizophrenia more logical than healthy volunteers?**
GARETH S. OWEN, JOHN CUTTING, and ANTHONY S. DAVID

**Correspondence:**

**Depression post-myocardial infarction**
A. Korszun, S. Stansfeld, and M. Frenneaux

**Depression post-myocardial infarction**
A. Shetty

**Depression post-myocardial infarction**
K. H. Kho

**Authors’ reply:**
P. de Jonge, J. P. van Melle, and J. Ormel

**Substance misuse disguised as ADHD?**
J. E. Berg
Authors’ reply

Heroin-assisted treatment: no difference in treatment retention
A. Al-Adwani and R. Nahata

Author’s reply
C. Haasen

Factors in those who repeatedly self-harm
M. Kripalani, R. Badanapuram, A. Gash, and S. Morris

Authors’ reply:
R. Young

Attachment disorders: an evolutionary perspective
H. P. O’Connell

Attachment disorders: an evolutionary perspective
H. Minnis, J. Reekie, D. Young, A. Gray, T. O’Connor, A. Ronald, and R. Plomin

One hundred years ago:

Stereotypy in Dementia Praecox [Étude clinique sur la stéréotypie des démêntia précoces]. (Arch. de Neurol., March, 1905.) Dromard, G.
Henry Rollin

Book reviews:

Fish’s Clinical Psychopathology: Signs and Symptoms in Psychiatry (3rd edn)
German E. Berrios

Cornelius Katona

Severe Dementia
Peter Bowie
Binge Britain: Alcohol and the National Response
Elizabeth Furlong

Dynamic Security: The Democratic Therapeutic Community in Prison
Christopher Cordess

Depression and Physical Illness
Ania Korszun

Formulation in Psychology and Psychotherapy: Making Sense of People’s Problems
Chris Mace

Medicine of the Person: Faith, Science and Values in Health Care Provision
Kedar Dwivedi

Reconceiving Schizophrenia
Matthew Broome

Love and Loss: The Roots of Grief and its Complications
Jan R. Oyebode

From the Editor's desk:

PETER TYRER

Correspondence:

Depression post-myocardial infarction
A. Korszun, S. Stansfeld, and M. Frenneaux

Depression post-myocardial infarction
A. Shetty

Depression post-myocardial infarction
K. H. Kho
Authors’ reply:
P. de Jonge, J. P. van Melle, and J. Ormel

Substance misuse disguised as ADHD?
J. E. Berg

Authors’ reply

Heroin-assisted treatment: no difference in treatment retention
A. Al-Adwani and R. Nahata

Author’s reply
C. Haasen

Factors in those who repeatedly self-harm
M. Kripalani, R. Badanapuram, A. Gash, and S. Morris

Authors’ reply:
R. Young

Attachment disorders: an evolutionary perspective
H. P. O’Connell

Attachment disorders: an evolutionary perspective
H. Minnis, J. Reekie, D. Young, A. Gray, T. O’Connor, A. Ronald, and R. Plomin
Highlights of this issue
BY SUKHWINDE S. SHERGILL

POSTNATAL DEPRESSION AND MODIFYING DISRUPTIVE BEHAVIOUR

Postnatal depression is likely to have an adverse effect on the mother–baby interaction. Poobalan and colleagues (pp. 378–386) reviewed the literature to examine the effects on the children of mothers treated for postnatal depression. They report that maternal treatment was associated with improvement in mother–baby interactions, but also had beneficial effects on the cognitive development of the children in certain studies. Disruptive behaviour during childhood is predictive of maladjustment during adolescence and adulthood. Could a preventive intervention at this early stage act to reduce subsequent criminal behaviour and improve academic performance? Boisjoli et al (pp. 415–419) suggest that such an intervention can indeed result in significant improvements in school performance and also have a beneficial effect by reducing the number of individuals acquiring a criminal record at follow-up. They suggest that such early preventive intervention for those at risk of antisocial behaviour may benefit these individuals in the longer term.

OPTIMAL TREATMENTS FOR ANOREXIA AND GENETIC RISK IN SCHIZOPHRENIA

There are limited contemporary high-quality data to guide the treatment of anorexia nervosa. Gowers et al (pp. 427–435) compared three treatment options: in-patient care, specialist out-patient treatment and generic treatment by the child and adolescent mental health team. The results show that all three groups improved over the 3 years of follow-up; however, only one-third of patients recovered fully. Contrary to their hypotheses, there were no significant differences between in-patient care and specialist out-patient care, when compared with routine treatment. They suggest that treatments administered without the patients’ cooperation may be of limited value. An economic evaluation of the project demonstrated that the specialist out-patient option may be the most cost-effective. There has been increased interest in the interaction between genetic vulnerability to developing schizophrenia and the contribution of cannabis and nicotine use. Zammit et al (pp. 402–407) found no association between two cannabinoid genes, or a nicotinic receptor gene, and schizophrenia or an interaction with drug use. Interestingly, they also failed to find a previously reported association between cannabis use and the catechol-O-methyl transferase (COMT) gene. They suggest that cannabis use is not associated with the ValMet COMT allele, and they do not find evidence of a gene–environment interaction between the COMT genotype and cannabis use exerting an effect on developing schizophrenia.

PTSD AND TMS TREATMENT FOR DEPRESSION

Post-traumatic stress disorder (PTSD) has been linked to disturbance of the hypothalamic–pituitary–adrenal axis. Meewise and colleagues (pp. 387–392) found that cortisol levels did not differ between patients with PTSD and healthy controls. However, they found lower levels of cortisol in those with PTSD who had been assessed in the afternoon, and lower levels in the control participants who had been exposed to trauma without developing PTSD. They suggest that the lower cortisol may be related to trauma exposure rather than PTSD per se. Repetitive transcranial magnetic stimulation (TMS) has been mooted as a treatment for depressive illness. Herwig et al (pp. 441–448) report the results of a study examining the efficacy of TMS as a strategy to augment standard antidepressant treatment. They found no differences in response rates between the TMS and control treatments.

SUPERVISED TREATMENT, DUAL DIAGNOSIS AND STRUCTURED COMMUNICATION

Compulsory community treatment has been proposed as a means of reducing the number of ‘revolving door’ admissions. Kisel & Campbell (pp. 373–374) report that this has not occurred where such community treatment has been introduced in Victoria, Australia. They suggest that this policy may require further review before being introduced wholesale in the UK, as currently proposed in the new Mental Health Act. Dual diagnosis is a common yet difficult problem in clinical practice, particularly where psychosis and substance misuse coexist. Johnson et al (pp. 451–452) found that training case managers in managing dual diagnosis did not result in the predicted reduction in patient admissions or in their drug and alcohol consumption. They suggest that a different approach, perhaps using specialist staff, may be required in these cases. While regular meetings with patients are a standard part of psychiatric treatment, they vary considerably in their format. Priebe and colleagues (pp. 420–426) implemented a standardised computer-mediated intervention focused on systematically assessing patients’ needs and their quality of life at each contact. Follow-up at 12 months demonstrated improvements in the patients’ quality of life and the reporting of fewer unmet needs and greater treatment satisfaction. They suggest that this could be more widely implemented at relatively low cost.
Psychiatry in pictures
EDITED BY ALLAN BEVERIDGE

Do you have an image, preferably accompanied by 100 to 200 words of explanatory text, that you think would be suitable for Psychiatry in Pictures? Submissions are very welcome and should be sent direct to Dr Allan Beveridge, Queen Margaret Hospital, Whitefield Road, Dunfermline, Fife KY12 0SU, UK.

Untitled (1889) by Charles Altamont Doyle (1832–1893)

Charles Altamont Doyle, the father of Arthur Conan Doyle, originally worked in an architectural post in Edinburgh until his heavy drinking rendered him unemployable. His family found a place for him in Blairerno House, a home for inebriates in the north of Scotland. However, he proved unmanageable and was committed to Montrose Asylum in 1885. He was judged by the admitting doctor to be 'very confused & bewildered'. He was to remain confused and the case notes repeatedly comment on his short-term memory problems. It is likely that Doyle was suffering from Korsakoff’s psychosis. Despite this, he continued to sketch and paint, demonstrating that creative work can co-exist with cognitive decline, at least for a time. This picture represents a fantastical portrayal of Montrose Asylum where ghostly horses and human figures float around its towers. Doyle’s Montrose artwork is wide-ranging and, as well as depicting scenes from asylum life, he painted giant birds, fairies, unicorns and exuberant vegetation. Some of the work he completed while an asylum inmate has been reproduced in Baker, M. (1978) The Doyle Diary (Paddington Press), from which the above picture was taken. Another example will feature in the December issue of the Journal.
Does compulsory or supervised community treatment reduce ‘revolving door’ care?

Legislation is inconsistent with recent evidence

STEPHEN KISELY and LESLIE ANNE CAMPBELL

Summary  Supervised community treatment to address ‘revolving door’ care is part of the new Mental Health Act in England and Wales. Two recent epidemiological studies in Australia (n>118 000), as well as a systematic review of all previous literature using appropriately matched or randomised controls (n=1108), suggest that it is unlikely to help.

Declaration of interest  None.

Although many patients have benefited from the de-institutionalisation of mental healthcare, there have been concerns that some have not received the care they require. Compulsory community treatment may help people stay in contact with services but remains controversial. Approaches include conditional discharge from hospital, community treatment orders for patients who are in the community, and court-ordered civil out-patient commitment.

One problem has been that much of the literature is based on opinion or uncontrolled studies. However, recent studies have used matching, multivariate analyses or randomisation to compare patients on compulsory community treatment with those not subject to such interventions. In the past year there have been four papers from two large studies based on the Victorian Psychiatric Case Register in Australia (n>118 000) (Burgess et al, 2006; Segal & Burgess, 2006a) as well as a systematic review of all previous literature using appropriately matched or randomised controls (n=1108) (Kisely et al, 2007). This is timely, as the Department of Health in England and Wales has included supervised community treatment in the new Mental Health Act to address the issue of ‘revolving door’ care (Department of Health, 2006).

DOES COMPULSORY COMMUNITY TREATMENT REDUCE ‘REVOLVING DOOR’ CARE?

The clearest indicator of whether compulsory community treatment helps ‘revolving door’ care is the number of bed-days rather than admissions. The intervention can only be the least restrictive alternative if individuals spend less time in hospital. In contrast, interpretation of the effect on admissions is less clear. Community treatment orders could conceivably either reduce admission rates, so allowing individuals to remain in their communities during treatment, or increase them, as a result of earlier identification of relapse.

ARE THERE OUTCOMES ON WHICH COMPULSORY COMMUNITY TREATMENT MIGHT HAVE AN EFFECT?

There are several potentially significant areas where the intervention was found to have an effect. Although community treatment orders used on initial discharge from hospital were associated with a higher risk of readmission, orders following subsequent admissions were associated with a lower risk (Burgess et al, 2006). However, we do not know the effect on bed-days, which may be the more critical measure of health service use. It was also difficult to determine whether this was also affected by changes in the use of compulsory community treatment over time, given that the number of orders increased from 919 in 1992 to 2260 in 2000 (Burgess et al, 2006).

Compulsory community treatment may also be more effective in early-episode cases when used within 30 days of initial admission to specialist services (Segal & Burgess, 2006b). However, the use of community treatment orders in first-episode cases would be impossible in most jurisdictions outside Australasia, where orders are limited to patients who have had substantial health service use in the year prior to the intervention. Another positive finding is that compulsory community treatment may reduce subsequent mortality (Segal & Burgess, 2006c). However, 10% of the patients in that study had dementia or other nervous system disease, which is not typical of populations elsewhere who are receiving compulsory community treatment, and patients with these diagnoses made up 29% of the total deaths.

WHAT ARE THE POLICY AND RESEARCH IMPLICATIONS?

None of the studies of compulsory community treatment is entirely satisfactory. The systematic review of the literature with appropriately matched or randomised controls that pre-dated the studies from Victoria was limited by the small number of studies (two RCTs and three CBA studies) (Kisely et al, 2007). Both RCTs were of court-ordered out-patient commitment in the USA, which may not be generalisable to other jurisdictions where compulsory
community treatment is initiated by clinicians and excludes patients with a history of violence (Swartz et al, 1999; Steadman et al, 2001). Of the three CBA papers, two were epidemiological studies from Western Australia which compared patients given community treatment orders with controls from within the same jurisdiction and internationally (n=652) (Preston et al, 2002; Kisely et al, 2003a). However, the two studies were restricted to patients given treatment orders in the first year of the legislation and may not reflect subsequent practice as clinicians gained experience in the use of the Act.

The two studies using the Victorian register were considerably larger and not subject to selection bias (Burgess et al, 2006; Segal & Burgess, 2006a). They also covered a decade’s experience of the legislation, and so may give a clearer picture of the longer-term effects than studies restricted to the first year of operation. However, there were also significant limitations. The authors did not match for date of placement on conditional release and so could not exclude the effect of other health-system changes that might have occurred between 1990 and 2000. In one study, conditional release and the outcome of interest had to occur in the same year (Burgess et al, 2006); in the other, the authors controlled for time of first contact with mental health services and mean year (Segal & Burgess, 2006a). Neither of these is quite the same as matching for discharge date. Controlling for time of first contact with mental health services could be affected by people arriving from other jurisdictions with pre-existing illness not captured by the Victorian Psychiatric Case Register. More importantly, although Segal & Burgess (2006a) controlled for time at risk, there was no stipulation that the event of interest (e.g. readmission or mortality) had to occur within a certain period of placement on conditional discharge. This means it could occur any time from 1 day to 10 years after the index date, whether someone was still on conditional discharge or not. Most previous work in this area has limited follow-up to 12 months after the order, as one has to be very cautious of ascribing an effect beyond a year following initial placement (Preston et al, 2002; Kisely et al, 2007).

In the case of mortality, the authors did not control for confounders such as lifestyle, psychotropic medication, reduced access to general medical care and the difficulties in recognising physical comorbidity in psychiatric patients with physical complaints (Kisely et al, 2005b; Segal & Burgess, 2006c). Furthermore, 72% of the deaths in people with mental health problems occurred in patients who had only ever been seen in primary care (Kisely et al, 2003b). Conditional discharge could therefore only play a very small part in addressing the increased mortality among patients with mental health problems, even if such a link were to be established.

Irrespective of how epidemiological studies have controlled for confounders, the selection of controls from the same jurisdiction as the community treatment order cases may be subject to confounding from variables such as social disability or characteristics of the treating team (Kisely et al, 2003a). These might explain why some patients and not others were given compulsory community treatment. Comparing jurisdictions with and without compulsory community treatment partially addresses this concern but raises the issue of comparability of the two health systems, especially with international comparisons (Kisely et al, 2003a).

In conclusion, there is limited evidence that compulsory community treatment will address the issue of the ‘revolving door’, at least in the short term, even though this is the Department of Health’s main justification for supervised community treatment in England and Wales (Department of Health, 2006). This issue illustrates how health policy remains determined by social or political factors as much as by evidence (Black, 2001). At the very least, researchers, funding bodies and policy makers should collaborate in evaluating the effects of the proposed legislation. Studies should ideally include a range of patient, family and health service outcomes using mixed methods, rather than focus on admission rates and lengths of stay. In the meantime, it might be more appropriate to acknowledge openly the limits of our knowledge, rather than rely on the illusion that evidence exists.

REFERENCES


The medical model is dead – long live the medical model

PREMAL SHAH and DEBORAH MOUNTAIN

Summary Many people criticise, and psychiatrists apologise, for the use of the ‘medical model’. We examine what is currently meant by this term and suggest a refinement of definition to reflect the ideals and contemporary practice of medicine. We propose that psychiatrists should use the medical model to improve and validate bio-psychosocial psychiatric medicine.

Declaration of interest None.

The term ‘medical model’ is frequently used in psychiatry with denigration, suggesting that its methods are paternalistic, inhumane and reductionist. This view has influenced mental health organisations, which in certain areas advocate a departure from the medical model, and contributes to the difficulties in leadership being played out between politicians, professionals and patients. The view has some support from within psychiatry (with some psychiatrists being apologists), from the 1960s’ anti-psychiatry movement, as well as from some in the recovery movement (Ralph et al., 2002). Although diversity is healthy, it may fuel unproductive rivalry to be recognised as the therapeutic agent between divergent therapies and agencies.

WHAT IS CURRENTLY MEANT BY THE MEDICAL MODEL?

There are various definitions. Clare (1980) suggested that it is a scientific process involving observation, description and differentiation, which moves from recognising and treating symptoms to identifying disease aetiologies and developing specific treatments. Wikipedia, the internet encyclopedia, currently defines it as ‘the predominant Western approach to illness, the body being a complex mechanism, with illness understood in terms of causation and remediation, in contrast to holistic, and social models’. The Disabled People’s Movement (http://www.bfi.org.uk/education/teaching/disability/thinking/medical.html) believes that it is based on a false notion of ‘normality’, with people being judged on what they cannot do. They believe that it sees people with disabilities as the problem, focusing on impairment, provoking fear and patronizing attitudes, the powerful doctor shutting the ‘disabled’ away. These definitions potentially combine to form the caricature of a reductionist, mechanistic, disability-enhancing approach, taken by powerful doctors towards patients.

Matters are aggravated in psychiatry because of the Descartian divide between biological and psychosocial psychiatry. Biological psychiatry is assumed to be mechanistic and reductionist, exclusively concerned with neuroimaging, genetics and medication. Psychosocial psychiatry, championed as being empowering, humane and holistic, is regarded as the antithesis and aligns itself to models such as Engel’s (1977).

IS BIOLOGICAL REALLY REDUCTIONISTIC?

The idea that the ‘biological’ is reductionist and undesirable leads to curious contradictions. The negative view of psychiatric drugs contrasts with views of drugs in other specialities or alternatives such as homoeopathy. The parallel assumption that psychosocial treatments are without risk, are holistic and the treatments of choice ignores evidence that some psychological treatments can cause damage (Rose et al., 2002). Furthermore, psychological treatments may work synergistically with drugs (Keller et al., 2000). In the extreme, advocating exclusive psychological approaches amounts to ‘psychological reductionism’ and could harm patients by denying them other effective treatments.

Biological explanations and treatments for diseases have helped to reduce fear, superstition and stigma, and to increase understanding, hope and humane methods of treatment (Tallis, 2004). For example, epilepsy is now better understood as a medical condition, which has reduced the perception of it being a fearful phenomenon of demonic possession. Logically a biological perspective in psychiatry should do the same.

Neuroscience demonstrates that biological, social and psychological experiences translate into changes in brain structure and function. Childhood sexual abuse (Teicher, 2000), personality trait differences (Breier et al., 1998), and psychological and pharmacological treatments (Seminowicz et al., 2004) have all been associated with differences in discrete brain systems, making it difficult to maintain the mind-body split, and offering a potential explanation for how bio-psychosocial treatments actually work.

We suggest that the difficulty lies in accepting that the human mind is also biological. This challenges cherished assumptions about our self-attributed uniqueness and the specialness of ‘the mind’. As Tallis (2004) comments, regarding one’s body as part of oneself but also objectively is difficult. How much more difficult is it to accept that what we experience as ourselves can be understood in terms of brain function? Banishing biology to reductionism merely defends our need to preserve the sanctity of ‘the mind’.

A CONTEMPORARY DEFINITION OF THE MEDICAL MODEL

We believe that we need a simple definition of the medical model, which incorporates medicine’s fundamental ideals, to facilitate clarity and precision, without denying its shortcomings. We propose that the ‘medical model’ is a process whereby, informed by the best available evidence, doctors advise on, coordinate or deliver interventions for health improvement. It can be summarily stated as ‘does it work?’

Face validity

Evidence has always been at the core of the ‘medical model’, encapsulated in Hippocrates’ dictum ‘first, do no harm’. This assumes that the doctor has specific knowledge and expertise (evidence) that
an intervention causes greater benefit than harm. Further, it is what most doctors do today, and it is what our patients expect – the days of treating on ‘gut feeling’ have long gone. Although some people question how much daily practice is evidence based (Imrie & Ramey, 2001), there is no call to abandon evidence and rely on faith or instinct alone. Practice is also increasingly determined by guidelines, with legal consequences for not doing so. We suggest that patients are not primarily concerned with a treatment’s ideological background but are more interested in what helps and what harms. Patients want us to provide a balanced view to enable them to decide.

**Ideology and assumption free**

The history of psychiatry well illustrates the perils of treatment by assumption or ideology. Our definition avoids invoking either ‘psychological’ or ‘biological’ ideologies. Although some people debate whether empiricism is reasonable, we contend that ideologies have fruitlessly divided pharmacology and psychotherapy (as well as psychotherapies themselves). How ‘it works’ is important but is secondary to defining ‘what works’. Not knowing how it works does not invalidate good evidence: vitamins were undiscovered when Lind conducted his scurvy trial (Bartholomew, 2002). Evidence, then, is the first critical step in elucidating what works and helps to validate our bio-psychosocial treatments, defining their dangers and efficacy.

**Scrutiny of interventions**

The model requires all interventions to be scrutinised using the same methodology. As each intervention is understood on its own merits, protagonists need to justify their treatment with standard evidence and not ideology. It is an unfair assumption that ‘mind’ treatments are beyond scrutiny and that some treatments are somehow innately ‘better’ than others. Such scrutiny is particularly important when public money funds interventions (e.g. in the National Health Service). Indeed, the benefit of the medical model is that it justifies expanding non-pharmacological as well as drug treatments.

**THE MEDICAL MODEL AND POWER**

We do not believe that the medical model is simply about doctors’ apparent power. The relationship between a patient and their doctor is complex. Patients initially seek a doctor because they believe this may be useful, which could be seen as conferring power to the doctor. However, getting better has always been an active process involving seeking help, evaluating options and making decisions about treatments. Patients do choose not to engage with treatment (e.g. only half of patients collect their prescribed medications), and may sabotage their own care. When vulnerable, people respond in various ways which are probably influenced by their experience of life: some feel powerless whereas others are motivated. Patient behaviours and expectations are not passive – if they were, patients would meekly stop smoking, curb their alcohol intake and have cervical smears! The doctor’s task then is to advise on the most effective intervention, the patient’s task being to decide and act on that advice while making sense of complex and conflicting emotions. ‘Patient empowerment’, therefore, has little to do with rescuing patients from the medical model. In this context, Engel’s psychosocial model does not contradict the ‘medical model’ but rather enhances it.

**FUTURE OF THE MEDICAL MODEL IN PSYCHIATRY**

We believe that it is important to adhere to the medical model as we have defined it (‘does it work?’). If we do not we may lose our hard-earned gains in defining effective psychiatric treatments. We must continue to gather quality evidence in order to establish which interventions work. This may be particularly challenging when defining which elements of psychosocial treatments are effective. This is critical to improve the credibility of psychiatry as a medical specialty, which has been attacked because psychiatric treatments are multi-dimensional.

We also contend that the well-informed psychiatrist who uses the medical model is ideally positioned to challenge those who engage on both sides of Descartian extremism. We need to acknowledge that our medical approach may sit uncomfortably with other doctors and mental health professionals, who may not perceive that psychosocial factors and interventions translate into biology. However, we should use and encourage the developments in neuroscience, which blur the distinction between mind and brain and which may demonstrate how bio-psychosocial interventions actually work. This offers the potential to bring rationality, specificity and validity to our interventions.

Most importantly, we should not apologise for using the medical model. Instead, we should challenge those who use it as a professional attack and question what is being criticised. We should not believe that the medical model is only about doctors’ powers, but remind ourselves that patients are active participants in the interaction. Medicine has always been about helping patients ‘take charge’ of their recovery by whatever means available.

Finally, we should rigorously challenge those who regard the psychiatrist as a uni-dimensional pharmacologist and reductionist. Psychiatrists’ training involves the use of biological, social and psychological treatments, a fact recognised in statute (e.g. the Mental Health (Care and Treatment) (Scotland) Act 2003). Thus, we should highlight what is undoubtedly psychiatry’s best asset (an area that the public may accuse our specialist colleagues of lacking): that of being a medical specialty in which the specialist understands and uses the holistic bio-psychosocial approach.

**REFERENCES**


Effects of treating postnatal depression on mother–infant interaction and child development

Systematic review


**Background**
Postnatal depression has detrimental effects on the child’s cognitive and emotional development.

**Aims**
To assess the benefits of treating postnatal depression for mother–infant interaction and child development.

**Method**
A systematic search was made of 12 electronic bibliographic databases for randomised controlled trials and controlled clinical trials on treatment of mothers with postnatal depression, where outcomes were assessed in children; findings were assessed.

**Results**
Only eight trials met the inclusion criteria. Of those included, interventions varied widely but all involved therapies directed at the mother–infant relationship. One study with intensive and prolonged therapy showed cognitive improvement, whereas two others with briefer interventions improved maternal–infant relationships but did not affect the child’s cognitive or behavioural development. All five studies assessing only mother–infant relationships showed improvements.

**Conclusions**
Cognitive development in children of depressed mothers, along with better mother–infant relationships, might be improved with sustained interventions. Trials assessing treatments for postnatal depression would benefit from looking more closely at benefits for children as well as mothers, using validated objective measures.

**Declaration of interest**
None.

Postnatal depression affects a mother’s ability to cope with the care of her baby, and limits her capacity to engage positively with the baby in social interactions (Murray et al, 2003a). Many studies report the detrimental effects of postnatal depression on the cognitive and emotional development of children (Sharp et al, 1995; Murray & Cooper, 1996; Murray et al, 1999; Hay et al, 2003) and so it appears that this disorder has adverse consequences for two generations of individuals. Reported prevalence of postnatal depression varies from 10% to 22% (O’Hara & Swain, 1996; Josefsson et al, 2001; Coates et al, 2004), with about 70,000 women experiencing postnatal depression in the UK alone every year (Glover et al, 2002). Therefore, with its high prevalence and transgenerational effects, this disorder constitutes a highly significant public health problem.

Interventions designed to prevent postnatal depression in mothers have been disappointing (Ogrodniczuk & Piper, 2003); Austin, 2004; Brockington, 2004; Dennis, 2005), whereas studies focusing on the early detection and treatment of postnatal depression have shown positive effects. However, these studies have concentrated on benefits to mothers (Appleye et al, 1997; Misri et al, 2000; Cooper et al, 2003) and it is unclear whether these interventions benefit child behaviour or development. This cannot be assumed, because the relationship between maternal depression and adverse child behavioural outcomes may also be genetically mediated, and not just the result of a causal effect of maternal behaviour on child functioning (Kim-Cohen et al, 2005). Nevertheless, the most recent Scottish guidelines published in 2002 (Scottish Intercollegiate Guideline Network, 2002) recommended that future research should assess the effects of interventions not just on mothers but also on the whole family unit. It is therefore important, for both research and clinical practice, to examine the effects of treating postnatal depression on the cognitive and psychosocial development of children. Therefore we sought to examine this issue within the context of a systematic review.

**METHOD**

**Literature search**
A systematic search was undertaken of 12 electronic bibliographic databases – MEDLINE, EMBASE, CINAHL, the Cochrane Library (Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Central Register of Controlled Trials and the American College of Physicians Journal Club), the Health Technology Assessment database, EBSCO, Zetoc, Applied Social Science Index and Abstracts, PsyCINFO, the Social Sciences Citation Index, the British Nursing Index and the Allied and Complementary Medicine database – covering the period 1966 to 2005. Search words used in Medline were post partum depression, postnatal depression, maternal depression, perinatal disorder of melancholy, postpartum OR postnatal OR psychiatric disorders. These were combined with terms for trials and children such as Randomised controlled trial, controlled clinical trial, clinical trials, random allocation, double OR single blind methods, evaluation studies, cross over studies AND children, infant, toddler, baby, adolescents. No language restriction was applied. The Medline search strategy was modified for the search of other databases. Reference lists of all included articles and review articles were also checked to identify any other relevant studies. Owing to time constraints, ‘grey’ literature such as conference proceedings and unpublished data were not searched and authors reporting on maternal outcomes only were not contacted for queries on measurements of child outcomes.

**Inclusion criteria**
Studies that met the following criteria were selected: randomised controlled trials and controlled clinical trials; all types of treatment interventions (pharmacological and non-pharmacological) for mothers diagnosed with post-partum depression; and outcomes measured in children up to 14 years of age. Studies that measured outcomes in other siblings in the family were also included in the review.
Exclusion criteria
Studies were excluded if they reported non-randomised interventions; interventions only for maternity 'blues' or maternal psychosis; or preventive interventions during the antenatal or early postnatal period, for participants identified to be at risk.

Assessment of studies
All identified abstracts were scanned by two reviewers independently. Each article that met the inclusion criteria was critically appraised, using a standard data extraction form, independently by two reviewers. Any queries about inclusion were discussed with the second reviewer and any disagreement was resolved by discussion with referral to a third reviewer. Information on study design, setting, sample characteristics, intervention details, measurement instruments used and follow-up was extracted. Child behaviour, mother–infant interaction or mother–infant relationship, and infant or child cognitive development (such as communicative, attentional and social skills) were the main outcomes assessed.

Quality assessment
The methodological quality of each included study was also assessed using a standard quality assessment form adapted from the Cochrane Collaboration and the Jadad scale (Jadad et al., 1996). Primary studies were assessed on the quality of random allocation of concealment; comparability of groups at baseline; masking of healthcare providers; outcome assessor's masking to intervention; time of follow-up, and percentage followed up; details of those leaving the trial; validation of the outcome measures used; reporting of outcomes (self-reported or objective measurement); and intention-to-treat analysis. Each of these criteria was graded from 0 to 2 according to the strength of compliance, giving a maximum total of 20. Each study was subsequently classified on the basis of the score obtained, with total scores below 10 considered to be weak, scores of 10–15 considered as moderate and scores above 15 as strong in quality.

Data analysis
Combination of results using meta-analysis was inappropriate owing to the heterogeneity of the included studies, which were varied in their interventions, outcome measurements and target populations. However, comparisons across studies were made, and direction of effect size discussed. The results are summarised according to the child outcome measures assessed.

RESULTS
The search identified a total of 4362 abstracts, which were scanned and full texts of 147 potentially eligible articles were critically appraised. Eight randomised trials or controlled clinical trials (nine papers) that met the inclusion criteria assessing the effects of treatment of postnatal depression on child outcomes were included in this review. The selection process is summarised in Fig. 1.

Of the eight studies identified, only three measured cognitive development in children (Cicchetti et al., 2000; Clark et al., 2003; Murray et al., 2003b) and the other five studies (six papers) assessed the effects of treatment on the mother–infant interaction or relationships (Meager & Milgrom, 1996; Hart et al., 1998; O’Hara et al., 2000; Horowitz et al., 2001; Onozawa et al., 2001; Glover et al., 2002). We considered that this was still an important outcome measure, in that the quality of the early mother–infant relationship is likely to have important consequences for later emotional and behavioural development (Murray et al., 1999). Individual studies are presented according to the child outcomes measured.

Cognitive development in children
The three studies that assessed cognitive development in children (Table 1) were varied in their participants and the interventions. Murray et al. (2003b) had only mothers as participants, whereas the other two studies (Cicchetti et al., 2000; Clark et al., 2003) included mothers and their infants. Cicchetti et al. (2000) investigated the efficacy of toddler parent psychotherapy (n=43) in postnatally depressed women, comparing them with 54 depressed mother–infant pairs not receiving this therapy 61 normal mother–infant pairs with no intervention (Table 1). The intervention lasted over a year (weekly for 57 weeks). The Bayley Scales of Infant Development (BSID) were used at baseline and the Wechsler Preschool and Primary Scales of Intelligence (WPPSI–R), which measure IQ in children aged 3–7 years, were used for older children. At the end of the trial, the WPPSI–R full-scale IQ score showed statistically significant differences between the groups (P=0.008). Whereas, the infants from the intervention group had IQ scores as high as those of the non-depressed control group (P=0.9), the infants in the control group with untreated depression had significantly lower full-scale IQ scores than either of the other two groups (non-depressed control group, P=0.02; intervention group, P=0.02). Further analysis demonstrated that these overall differences were attributed to group differences in the verbal performance IQ scales of the children (P=0.02; mean verbal IQ scores were 104.2 in the intervention group, 103.7 in the non-depressed control group and 97.5 in the depressed control group), with no statistically significant difference between the groups for performance IQ (P=0.10). Analyses of the differences between the
<table>
<thead>
<tr>
<th>Study</th>
<th>Location, design and quality of study</th>
<th>Details of participants</th>
<th>Intervention and delivery</th>
<th>Outcome measurements and results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cicchetti et al. (2000) USA</td>
<td>Design: RCT, Quality of study: moderate</td>
<td>Mothers and infants (n=187)</td>
<td>Toddler parent psychotherapy (n=43) compared with depressed (controls) (n=54) and non-depressed controls (n=61). Psychotherapy was delivered on a weekly basis for an average of 57 weeks between child ages 20 months and 3 years, mean number of sessions was 45. No follow-up.</td>
<td>Primary outcome: child cognitive development, assessed by the Bayley Mental Development Index in infants and Wechsler Preschool and Primary Scales of Intelligence in older children. Results: significant differences in favour of the psychotherapy group (n=43) compared with depressed controls (n=54) for cognitive development in children. No significant difference between the psychotherapy group and non-depressed controls (n=61).</td>
</tr>
<tr>
<td>Murray et al. (2003) UK</td>
<td>Design: RCT, Quality of study: strong</td>
<td>Mothers only (n=193)</td>
<td>Routine primary care (n=52) compared with one of the three index interventions such as counselling (n=48), cognitive-behavioural therapy (n=43) and psychodynamic therapy (n=50). Therapy was conducted on a weekly basis, from 8 weeks to 18 weeks post-partum, by specialists in the three intervention fields; routine care was delivered by GPs and health visitors. Follow-up at 18 months and 5 years.</td>
<td>Primary child outcomes: child behaviour; mother-infant relationship, infant attachment and child cognitive development, various tools were used to assess child outcomes at different time points. Results: all the treatments showed some short-term benefit in mother-infant relationship and child outcomes with counselling and cognitive-behavioural therapy more beneficial than psychodynamic therapy. In the long term no significant benefit was observed with any of the treatments for all outcomes.</td>
</tr>
<tr>
<td>Clark et al. (2003) USA</td>
<td>Design: CCT, Quality of study: moderate</td>
<td>Mothers and infants (n=39)</td>
<td>Two active intervention groups compared with a waiting-list control group (n=11). The active interventions were mother-infant therapy (n=13) and interpersonal psychotherapy (n=15) delivered by psychologists and social workers. No follow-up.</td>
<td>Primary child outcomes: mother-infant interaction and child cognitive development. Child domains of the Parenting Stress Index were used to measure the stress related to the child domain. The Parent-Child Early Relational assessment was used to assess the quality of the mother-child relationship. Bayley Scales of Infant Development - mental scales were used to assess the cognitive and motor development of the infants and toddlers. Results: mother-infant therapy (n=13) and interpersonal psychotherapy (n=15) were equally effective in terms of mother-infant relationship compared with the waiting-list control group (n=11) but no statistically significant difference was found between the three groups in child cognitive development.</td>
</tr>
</tbody>
</table>

CCT, controlled clinical trial; GP, general practitioner; RCT, randomised controlled trial.
Table 2  Studies that assessed mother—infant relationships

<table>
<thead>
<tr>
<th>Study</th>
<th>Location, design and quality of study</th>
<th>Details of participants</th>
<th>Intervention and delivery</th>
<th>Outcome measurements and results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meager &amp; Milgrom</td>
<td>Australia</td>
<td>Mothers only (n=20)</td>
<td>A 10-week post-partum support programme (n=10) was compared with a waiting-list control group (n=10). Each session was 1.5 h in duration and conducted by clinical psychologist No follow-up</td>
<td>Primary outcomes: maternal&lt;br&gt;Secondary child outcome: Parenting Stress Index – child domain&lt;br&gt;Results: no statistically significant difference between support group (n=6) and controls (n=6) in PSI. Child domain sub-scale showed statistically marginal deterioration in control group (P=0.05) between baseline and end of trial</td>
</tr>
<tr>
<td>O’Hara et al (2000)</td>
<td>USA</td>
<td>Mothers only (n=120)</td>
<td>Interpersonal psychotherapy (n=48) compared with a waiting-list control group (n=51) over 12 weeks Interpersonal psychotherapy was administered in 12-hour-long individual sessions by psychotherapist with training in this intervention No follow-up</td>
<td>Primary outcomes: maternal&lt;br&gt;Secondary outcome: mother—infant relationship, assessed using the self-reported Social Adjustment Scale, which looks at relationship with children older than 2 years, and the Post Partum Adjustment Questionnaire, which looks at relationships with children other than the baby&lt;br&gt;Results: statistically significant difference between the groups favouring the interpersonal psychotherapy (n=48) in relationship with children — 2 years old and children other than the baby, but no significant difference in relationship with the new baby</td>
</tr>
<tr>
<td>Hart et al (1998)</td>
<td>USA</td>
<td>Mothers and infants (n=27 dyads)</td>
<td>Teaching mothers to use Mother’s Assessment of the Behaviour of her Infant (n=14 dyads) compared with a weekly written report of infants behaviour by control group mothers (n=13 dyads) No follow-up</td>
<td>Primary outcome: Neonatal Behavioural Assessment Scale, designed to evaluate behavioural and neurological functioning in neonates and young infants&lt;br&gt;Results: Ratings by examiners showed significant improvement in social interaction and state organisation in treatment group, (where mothers administered MABI periodically at home; n=14 dyads) compared with the control group (no MABI administered by mothers but had a written report; n=13 dyads)</td>
</tr>
<tr>
<td>Horowitz et al (2001)</td>
<td>USA</td>
<td>Mothers and infants (n=122)</td>
<td>Home visits only (n=57) compared with home visits and interaction coaching for at-risk parents and infants (n=60) Three home visits over 18 weeks by research nurse, and interaction coaching delivered by advanced practice nurses and research assistants. Each session lasted 15 min No follow-up</td>
<td>Primary outcome: mother—infant responsiveness, assessed with the Dyadic Mutuality Code&lt;br&gt;Results: statistically significant difference in favour of interaction coaching group (n=60) compared with group with only home visits (n=57) for increase in responsiveness (P=0.006); this was maintained over time (P=0.025)</td>
</tr>
</tbody>
</table>

*Continued*
<table>
<thead>
<tr>
<th>Study</th>
<th>Location, design and quality of study</th>
<th>Details of participants</th>
<th>Intervention and delivery</th>
<th>Outcome measurements and results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onozawa et al (2001) UK</td>
<td>Design: RCT</td>
<td>Mothers and infants (n=34)</td>
<td>Support group only (n=12) as controls compared with infant massage in addition to support group (n=13)</td>
<td>Primary outcome: mother–infant interaction, assessed by video recording and rated according to global rating for mother–infant interactions at 2 months. This rates the maternal contribution, infant’s contribution and the interaction itself. Results: statistically significant difference in favour of infant massage (n=10) compared with the support group (n=12), with marked improvement in mother–infant interaction (P = 0.0004).</td>
</tr>
<tr>
<td></td>
<td>Quality of study: moderate</td>
<td>Age: mothers, median 18–45 years; infants, 9 weeks</td>
<td>Support group consisted of 30 min per week informal group discussion for 5 weeks. Infant massage consisted of a short period of relaxation followed by infant massage for 1 h by trained instructors, in addition to support group for 5 weeks</td>
<td>No follow-up</td>
</tr>
<tr>
<td>Glover et al (2002)† UK</td>
<td>Design: RCT</td>
<td>Mothers and infants (n=34)</td>
<td>Support group only (n=12) as controls compared with infant massage in addition to support group (n=13)</td>
<td>Primary Outcome: mother–Infant interaction, assessed by the video recording and rated as above. Results: statistically significant improvement in mother–infant interaction was achieved for the mothers who attended the massage class (n=12) compared with the control group (n=12) over time – from baseline to the final session (P &lt; 0.001).</td>
</tr>
<tr>
<td></td>
<td>Quality of study: moderate</td>
<td>Age: mothers, median 18–45 years; infants, median 9 weeks</td>
<td>Intervention described above (Onozawa et al, 2001)</td>
<td></td>
</tr>
</tbody>
</table>

MABI, Mother’s Assessment of the Behavior of her Infant; PSI, Parenting Stress Index; RCT, randomised controlled trial.
† The same study is reported by Onozawa et al (2001) and Glover et al (2002), but the results are presented slightly differently.
standardised full-scale IQ and standardised Bayley Mental Development Index (MDI) score also revealed significant differences between groups (P = 0.05). Following the same pattern, post hoc tests found that infants in the depressed control group were significantly less well developed than those in the other two groups, who did not differ from each other. Additional factors that could potentially account for the intervention effects were used as additional covariates in the analysis of covariance (ANCOVA) to examine cognitive functioning, controlling for the baseline MDI score. These factors included maternal education, subsequent depressive episodes, marital status, mother’s work status and presence or absence of comorbidities. In this analysis only subsequent depressive episodes accounted for the cognitive outcome difference at 36 months, with significant differences for full-scale IQ scores (P = 0.012) and verbal scores (P = 0.015) but not for performance IQ scores. This followed the same pattern with worse outcomes for the depressed control group compared with the other two groups.

Murray et al (2003b) evaluated the effect of three types of psychological treatment – non-directive supportive counselling, cognitive–behavioural therapy and brief psychodynamic psychotherapy – compared with routine primary care on the mother–child relationship, and continued to measure child outcomes up to 5 years of age (Table 1). The results were adjusted for relationship and behavioural problems prior to treatment, and also for social adversity. At the end of the treatment period (after 4 months), all three treatments significantly improved the quality of the mother–infant relationship but had no effect on the level of behavioural management problems. At 18 months follow-up, Behavioral Screening Questionnaire (BSQ) scores showed a significant difference between the groups (Kruskal–Wallis test = 9.04, P = 0.03), reflecting greater effects of active treatment compared with routine care. Using a general linear model assuming a gamma distribution (skewed distribution) rather than the usual normal distribution for the BSQ scores, analysis of the treatment groups data showed significant differences for the non-directive counselling group (χ² = 12.19, P = 0.001), the brief psychodynamic therapy group (χ² = 4.06, P = 0.03) and the cognitive–behavioural therapy group (χ² = 3.52, P = 0.06) when compared with the control group. However, scores on infant attachment and child cognitive development were similar in the four groups (Kruskal–Wallis = 0.78, P = 0.85). At the end of 5 years, child emotional and behavioural difficulties were assessed using maternal reports on the Rutter A2 Parent Scale for Pre-school Children and teacher reports on the Pre-school Behavior Checklist. The differences between the four groups on the Rutter A2 scale did not quite reach significance (Kruskal–Wallis = 7.19, P = 0.07). Scores on teachers’ reports of child behaviour difficulties (Kruskal–Wallis = 0.10, P = 0.99) and measures of cognitive development using the McCarthy scales (Kruskal–Wallis = 0.53; P = 0.91) more clearly showed an absence of differences between the groups. In addition, no significant treatment effect was observed even after controlling for level of social adversity in the linear model.

Clark et al (2003) used a short-term intervention that compared a mother–infant therapy group (n = 13) with interpersonal psychotherapy (n = 15) and a waiting-list control group (n = 11). They examined the effects of treatment on the mother–infant relationship and child cognitive and motor development but did not examine behavioural symptoms or long-term effects. Using ANCOVA with pre-treatment scores and maternal age as covariates, the child domain scores of the Parenting Stress Index were only significant for the ‘child adaptability’ (P = 0.036) and ‘reinforces parent’ (P < 0.001) domains. Post hoc tests on both of these domains showed the difference was an improvement for the active intervention groups compared with the waiting-list control group, with no statistically significant difference between the two active interventions. Similarly, Parent–Child Early Relational Assessment ratings showed statistically significant group differences for factor 1 (maternal positive affective involvement and verbalisation; P = 0.005) and factor 2 (maternal negative effect and behaviour; P = 0.035). For factor 1, both of the therapy groups scored higher than the waiting-list group, demonstrating more maternal positive affective involvement and verbalisation with their infants; again, the two intervention groups did not differ from each other. However, for factor 2 the mother–infant therapy group was significantly different from the waiting-list group whereas the interpersonal therapy group was not. Although values were not reported in the paper, its authors commented that no statistically significant difference was found among the three groups for the mental scales of the BSID.

Because of the heterogeneity of these three studies meta-analysis was considered inappropriate, but some qualitative comparison may be worthwhile. Clark et al (2003) had a small sample size and a short-term intervention with no follow-up. Cicchetti et al (2000) had a much more intensive and prolonged intervention which showed beneficial effects on cognitive development in children. However, there was no long-term follow-up similar to that by Murray et al (2003b) to assess the sustainability of the benefits. On the other hand, Murray et al (2003b), with comparable sample sizes to the Cicchetti study but a much shorter duration of therapy, showed some short-term benefits from all the treatments for the mother–infant relationship and early child outcome, but did not show any long-term benefit in emotional and behavioural adjustment or cognitive development after 5 years of follow-up.

Mother–infant interaction or relationship

Five studies assessed only mother–infant relationships (Table 2). Once again the interventions varied, ranging from a support group for mothers with depression to interpersonal psychotherapy, making any formal pooling of the results inappropriate.

Two of the studies (Meager & Milgrom, 1996; O’Hara et al, 2000) had only mothers with postnatal depression as participants. Meager & Milgrom (1996) observed changes in the children but had small sample sizes (10 in each arm). They measured scores on the Parenting Stress Index which did not change significantly over time for either group. Analysis of variance (ANOVA) was used to assess changes in child domains and to explore the variation between ‘mothers’ and ‘time spent by mothers in support programme’ over all 10-week time points. This did show a deterioration in the child domain subscale scores, with a post hoc least significant difference over time of 13.42 (P = 0.05) for the control group but not for the intervention group. Although the support group showed marginal benefits, this was based on a sub-scale reflecting the mother’s perception of the infant rather than the child outcome itself. In the other study with only mothers as participants, O’Hara et al (2000) assessed the effect of
interpersonal psychotherapy; this was a larger study but conducted over only two weeks. The results, repeated ANOVA measures, favoured interpersonal psychotherapy over control for two of the sub-scales of the Social Adjustment Scale and the Post-partum Adjustment Questionnaire (PPAQ) reflecting the quality of the parent-child relationship. These sub-scales were ‘relationship with older children more than 2 years’ ($P < 0.05$) and ‘mother’s relationships with children other than the baby’ ($P = 0.005$). However, the PPAQ showed no significant difference between the two groups with respect to the ‘relationship with the new baby’ sub-scale ($P = 0.13$).

Three studies (Hart et al, 1998; Horowitz et al, 2001; Onozawa et al, 2001) had both mothers and their infants as participants, but once again interventions varied. Horowitz et al (2001) studied the effects of ‘interactive coaching’ on the mother-infant relationship. Here the coached group had a statistically significant higher Dyadic Mutuality Code mean score than the control group at 10–14 weeks ($P = 0.002$) and at 14–18 weeks ($P = 0.029$). Responsiveness between mother and the infant using repeated measures of ANOVA also showed a significant difference between the treatment and control groups ($P = 0.006$) and over time ($P = 0.025$). Thus, the increase in responsiveness that occurred following the intervention was maintained at least to 18 weeks. A study reported on by both Onozawa et al (2001) and Glover et al (2002) compared infant massage with a support group. Mother-infant interactions were assessed by video recording and rated as ‘maternal contribution to the interaction’, ‘infant’s contribution’ and the mother-infant interaction itself. Onozawa et al (2001) found an improvement in the maternal-infant interaction in the massage group compared with the support group. Glover et al (2002) reported the same results slightly differently, presenting the mother–baby interaction scores over time, showing that for mothers who attended the massage class a statistically significant improvement was achieved ($P < 0.001$) compared with the control group.

Hart et al (1998) reported a different intervention in which the researchers sought to assess whether training depressed mothers to examine their infants might improve their child’s developmental outcomes. Those in the intervention group (14 mother and infant pairs) observed an administration of the Neonatal Behavioral Assessment Scale (NBAS) soon after the delivery of the baby, by a trained examiner. The examiner explained the significance of various infant behaviours, such as turning toward a sound source and tuning out distractions. Mothers were then given feedback on their infant’s behaviour, and given the opportunity to discuss this. After the administration of NBAS by examiners, mothers were taught to administer a similar instrument, the Mother’s Assessment of the Behavior of her Infant (MABI), independently. They were then instructed to repeat the administration at home at 1-week intervals for 1 month. For the control group (13 mother and infant pairs), mothers were not present when the NBAS was administered by the examiners at the delivery, although they were asked to periodically complete written assessments at home of their parenting attitudes and the infant’s development. Outcomes consisted of both examiner and maternal ratings on the NBAS at the end of the trial. Ratings of infants by examiners (unaware of the mother’s group status) revealed that after 1 month, infants in the experimental group (where mothers administered the MABI periodically at home) were performing better than the infants in the control group for social interaction ($P < 0.05$) and state organisation ($P = 0.05$). Mothers in both groups, although not significantly different from each other, rated their infants as showing significant improvements over time for social, motor and state organisation indicative of developmental progress. The authors concluded that NBAS/MABI enhanced social interaction and state organisation in children, even though mothers’ perceptions of their infant’s behaviour were not different between the two groups.

Overall, these five studies measuring the mother-infant relationship showed improvement irrespective of the type of intervention and the target population (either mothers only, or infants along with mothers). However, the instruments used to measure outcome in these studies need to be taken into consideration. The Parenting Stress Index used by Meager & Milgrom (1996), measured parental adjustment to parenting rather than outcomes in children per se. In the study by Hart et al (1998), the measure was the same as the intervention that had been used more often in the active treatment group. It is therefore unclear whether this study simply detected the effects of the infants practising the assessment rather than any genuine therapeutic effect.

**DISCUSSION**

With many women experiencing postnatal depression and the apparent ineffectiveness of preventive strategies for postnatal depression in women at high risk of this disorder, early detection and treatment are increasingly being recognised as the priority (Ogrodniczuk & Piper, 2003; Dennis, 2005). Despite the wide range of treatments (including antidepressants, progesterone, cognitive–behavioural therapy and interpersonal therapy) that are available and beneficial for mothers, an important question is whether treatment of postnatal depression has demonstrable benefits for children, given the significant short-term and long-term consequences of the disorder on both the baby and its siblings.

This review is the first to search systematically for clinical randomised controlled trials that assessed the effects of treatment of postnatal depression on the physical and mental health of children rather than that of the mothers. We were also interested to find out whether this depended upon the type of treatment and if it was influenced by maternal variables such as persisting depression or psychosocial adversity.

**Methodological issues and limitations**

Despite a comprehensive search, only eight studies (nine papers) assessing child outcomes in response to postnatal depression treatment were identified, since most studies focused only on maternal outcomes. Treatment interventions in the identified studies varied widely, but all contained elements that sought to influence child development or the mother–child relationship. This highlights a potentially important issue in relation to the aetiology of postnatal depression, in that it may be driven by maternal difficulties in forming a relationship with the infant. The interventions described in this review may serve to treat depression by addressing these underlying relationship conflicts directly, or through simple effects on the depression with naturally positive consequences for the relationship. Most studies were not able to address this distinction. Improved infant behaviour could have been a direct reflection of improvement in maternal mood. This is particularly likely for some of the outcome measurement tools used, such as the Parenting Stress Index (Meager & Milgrom,
Although patient numbers might have been
therapies and therapists’ skill levels.
emerged from our review concerned group
prenatal stage or interventions directed at
view did not cover interventions in the
the mother–infant relationship. Further, our
review did not cover interventions in the
between treatments can be confounded by dif-
Consequently, potential differences be-
which showed no specialist therapist effect.
relationship issues and involves infants in treat-
directly addresses mother–infant relation-
relationship had actually improved.
attitudes and perceptions. It seems likely
measure child behaviour directly but were
1996; Clark et al, 2003) which did not
self-report measures, reflecting maternal
attitudes and perceptions. It seems likely
that scores on these measures would
improve with better mood, irrespective of
whether the child’s behaviour or the
relationship had actually improved.
Nevertheless, for studies that did
measure child outcomes directly, it was
impossible to disentangle whether improve-
ment of maternal–infant relationship would
result from simple treatment of the post-
natal depression, or whether this improve-
ment is dependent on a treatment that
directly addresses mother–infant relation-
ship issues and involves infants in treat-
ment. One possible pointer is that the
interventions by Cicchetti et al (2000) and
Murray et al (2003b) are likely to have
been of equal benefit in the treatment of de-
pression; this suggests that the specific ben-
etits found in the study by Cicchetti et al
(2000) stemmed from positive influences on
the mother–infant relationship. Further, our
review did not cover interventions in the
prenatal stage or interventions directed at
mothers who were not clinically depressed.
A final methodological point that
emerged from our review concerned group
therapies and therapists’ skill levels.
Although patient numbers might have been
high, therapists tended to be few in number
and their experience varied; some were
students. Although usually training had been
provided to the study therapists (Clark et al,
2003), it is possible that outcomes re-
lected individual differences between
therapist skill levels rather than differences
between treatments unless it was controlled
for as in the study by Murray et al (2003b)
which showed no specialist therapist effect.
Consequently, potential differences be-
tween treatments can be confounded by dif-
ferences among those who are delivering
the treatments.

Findings of the review
Overall, we found that all treatments for
postnatally depressed mothers had some
benefits in improving the quality of the
mother–infant interaction and relationship,
the level of behavioural management
problems and cognitive development in
children. However, it is important to high-
light that observed improvements were
based on only a few studies, with very
different interventions and measurement
tools. Only the study by Cicchetti et al,
(2000), assessing toddler parent psycho-
therapy as an intervention, showed signifi-
cant improvement in the children’s
cognitive development. In this study the in-
tervention was much more intensive and
longer-lasting. The outcomes were mea-
sured objectively using standard validated
measuring tools, and the degree of improve-
ment following the intervention is likely to
have been clinically and developmentally
significant. However, this study did not fol-
low the children up to see if the benefits
were sustainable. In contrast, the long-term
follow-up study by Murray et al (2003b)
(5 years of follow-up) failed to show
sustainability of short-term benefits, but it
had an intervention of shorter duration
(only 18 weeks).

Improvements observed in the infants
could also have been a direct reflection of
improvement in maternal depression scores.
All the studies, irrespective of the
child outcome measured, made an assess-
ment of maternal depression scores and re-
ported improvement in the women’s
depression levels. Although no detailed
analysis was carried out for this review re-
garding the outcome in mothers, none of
the studies made any attempt to explore
an association between the improved ma-
ternal depression scores and the improved
infant outcomes. From this, there is not en-
ough evidence to show if the improvement
in maternal–infant relationship was a
consequence of improvement in maternal
depression alone or if there was an addi-
tional intervention or participant element
that could have improved the cognitive
status of the children.

Implications for practice
From the evidence available, treatment
interventions in mothers for postnatal
depression seem to have some benefits for
the mother–infant relationship. However, for
improving cognitive development in
children, in spite of one high-quality study
in our review providing strong evidence of
benefits, the long-term sustainability needs
to be assessed. Cognitive development in
children, along with mother–infant rela-
tionships, may also be best improved with
sustained interventions over a longer period.

Implications for research
In spite of the significant impact of post-
natal depression on children, most treat-
ment trials for this disorder treat mothers
in isolation and concentrate on maternal
outcomes. Even in studies with assessments
of both maternal and child outcomes, no
attempt had been made to determine
whether there was any association between
the maternal and child outcomes, which
could have considerable implications. A
well-conducted randomised controlled trial
with adequate power and long-term follow-
up, comparing two different potentially
effective interventions identified in this
review, is warranted to try to identify the
most effective intervention and assess the
impact of treatment of postnatal depression
on children. Furthermore, combining treat-
ments for maternal depression (such as
antidepressant medication) with therapies
focused on the mother–infant relationship
and investigating associations between im-
proved maternal depression and infant out-
comes may also be worthy of consideration.
Studies should also assess the effect of
treatment using directly rated child
measures rather than relying on maternal
self-reporting. Given that our review did
not cover interventions in the prenatal stage
or interventions directed at mothers who
were not clinically depressed, future re-
search should consider this group.

This review is unable to provide strong
evidence for a single effective intervention
to improve cognitive development in chil-
dren, given the disparate interventions.
The research does, however, suggest that
long-term, intensive interventions directed
at the mother–infant relationship may bring
about benefits in cognitive development of
the child. These potentially effective inter-
ventions should be further explored with
well-powered trials so that comparisons
can be made in order to achieve improve-
ment and sustainability over a long period.

ACKNOWLEDGEMENTS
The study was funded by the Scottish Evidence
Based Child Health Unit, National Health Service
Scotland.

REFERENCES
controlled study of fluoxetine and cognitive-behavioural
counselling in the treatment of postnatal depression.
BMJ, 314, 932–936.

intervention for perinatal distress, depression and
anxiety: where to from here? Archives of Women’s
Mental Health, 7, 1–6.


POOBALAN ET AL

AMUDHA S. POOBALAN, MBBS, MSc, Scottish Evidence Based Child Health Unit, Department of Child Health; LORNA S. AUCOT, TSc, PhD, Department of Public Health; LOUISE ROSS, BSc, MSc, Scottish Evidence Based Child Health Unit, Department of Child Health; W. CAIRNS, S. SMITH, MBChB, MD, MPH, MFPHM, PhD, FRCP, Department of Public Health; PETER J. HELMS, MBBS, PhD, JUSTIN H. G. WILLIAMS, MRCPsych, Department of Child Health, University of Aberdeen, Aberdeen, UK

Correspondence: Dr Amudha Poobalan, Department of Public Health, Medical School, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK. Tel: +44 (0)1224 555 934; fax: +44 (0)1224 550 925; email: a.poobalan@abdn.ac.uk

(First received 26 October 2006, final revision 26 April 2007, accepted 14 May 2007)
Cortisol and post-traumatic stress disorder in adults
Systematic review and meta-analysis*

MARIE-LOUISE MEEWISSE, JOHANNES B. REITSMA, GIEL-JAN DE VRIES, BERTHOLD P. R. GERSONS and MIRANDA OLFF

Background Post-traumatic stress disorder (PTSD) has inconsistently been associated with lower levels of cortisol.

Aims To compare basal cortisol levels in adults with current PTSD and in people without psychiatric disorder.

Method Systematic review and meta-analysis. Standardised mean differences (SMD) in basal cortisol levels were calculated and random-effects models using inverse variance weighting were applied.

Results Across 37 studies, 828 people with PTSD and 800 controls did not differ in cortisol levels (pooled SMD = −0.12, 95% CI = −0.32 to 0.080). Subgroup analyses revealed that studies assessing plasma or serum showed significantly lower levels in people with PTSD than in controls not exposed to trauma. Lower levels were also found in people with PTSD when females were included, in studies on physical or sexual abuse, and in afternoon samples.

Conclusions Low cortisol levels in PTSD are only found under certain conditions. Future research should elucidate whether low cortisol is related to gender or abuse and depends on the measurement methods used.

Declaration of interest None.

Exploring neuroendocrine function in patients with post-traumatic stress disorder (PTSD) may give insight into the pathogenesis of this stress-related disorder. One focus in the scientific literature has been on the hypothalamic-pituitary-adrenal (HPA) axis. Cortisol secretion is regulated by psychological stress to stimulate the HPA axis, adrenocorticotropic hormone (ACTH), which results in the production of glucocorticoids (cortisol) in the adrenal cortex. Cortisol serves to stop metabolic, neuronal defensive and immune reactions. Consequently, energy can be mobilised to cope adequately with the stressor. Studies using psychological stress to stimulate the HPA axis have shown an exaggerated cortisol response in PTSD (for review see de Kloet et al., 2006). However, under basal conditions PTSD generally – but inconsistently – has been associated with lower levels of cortisol. To elucidate this association we performed a systematic review of studies reporting basal cortisol levels in people with PTSD to investigate whether specific circumstances contribute to these inconsistencies.

METHOD
Identification of studies Relevant studies were identified by systematic searches of Medline, EMBASE, ScienceDirect and Psychlit for articles published between 1980 and March 2005. A free-text search was performed using the term posttraumatic stress disorder, OR PTSD AND cortisol, OR glucocorticoid. Reference lists of the articles obtained were checked for further relevant articles. We also considered inclusion of unpublished studies offered to us by researchers.

Inclusion criteria Studies were included when cortisol was measured in adults (aged 18 and above) with current PTSD and in controls with no current Axis I disorder (according to DSM–III, DSM–III-R, or DSM–IV criteria; American Psychiatric Association, 1980, 1987, 1994) or history of PTSD. Cortisol levels had to be determined with a standard biological assay. Furthermore, mean cortisol levels and standard deviations (s.d.) for both groups had to be described or had to be presented by the authors upon request. In the case of multiple papers from a single study, only the results of the publication with the highest number of participants was included.

Exclusion criteria Studies were excluded when: (a) focusing on a condition other than depression comorbid with PTSD (e.g. borderline personality disorder, cancer); (b) reporting on lifetime, 12-month diagnosis of PTSD or sub-threshold/partial PTSD; (c) cortisol level was measured within 1 month of trauma; (d) the HPA axis was pharmacologically challenged (e.g. by dexamethasone) before first cortisol measurement; (e) participants were anticipating any kind of social stressor, such as cognitive stress challenge, combat noise or personalised trauma script; (f) reporting in a language other than English. M.M and G.V. independently assessed each retrieved study and disagreements with respect to inclusion were resolved through discussion with M.O.

Data analysis For each individual study identified, we calculated the standardised mean difference (SMD) in cortisol levels between the PTSD and the control group and its associated variance. Hedges’ adjusted g was used to give a better estimate in cases of smaller sample sizes (Rosenthal et al., 1994). Random-effects models were fitted using inverse variance weighting to obtain pooled estimates of SMD and its corresponding 95% CI.

In all analyses, we used the SMD as our outcome measure to allow pooling across studies that used different types of measurement (i.e. urine, saliva, plasma, or serum) and to reduce the impact of measurement problems related to different sampling conditions.

Our first analysis included the data from all studies to obtain an overall pooled estimate and to examine whether there was heterogeneity in results between studies. Each study was included only once in this
overall analysis. If a study reported multiple types and times of measurements, the following hierarchy was used to select one measurement. Plasma samples were preferred above saliva, saliva above urine, and morning measurement above afternoon or evening. We used data from the earliest sample, or when measurements were related to time of awakening we selected the sample 30 min past awakening to be closest to the peak level of cortisol in the morning. For afternoon samples, we included the latest possible sample to compare circadian curves at the nadir of cortisol values. The $Q$-test was performed to examine whether there was more heterogeneity in the results than could be expected from chance alone. We also calculated the $I^2$ statistic, which expresses the percentage of total variation that can be attributed to heterogeneity rather than chance. Within this data-set, we also examined whether there were systematic differences between the types of measurement.

We then performed several specific subgroup analyses to examine whether the difference in cortisol levels between PTSD and control groups was influenced by other factors. Successive models were built to examine whether the SMD was significantly different in subgroups defined by a particular factor. The following factors were examined: time of measurement; gender; characteristics of the PTSD group, including type of trauma, years elapsed since trauma, presence or absence of comorbid depression; whether or not the control group was also exposed to trauma; and year of publication. Results of the subgroup analysis are presented as mean SMD together with 95% CI for each level of the factor. In addition, a formal test of interaction (e.g. whether the differences in SMD between the levels of the factor are zero) was performed to avoid overinterpretation of effects found in subgroups (Matthews & Altman, 1996; Altman & Bland, 2003).

The MIXED procedure in SAS version 9.1 for Windows was used to fit the various random-effects models as described by van Houwelingen et al (2002). $P$-values less than 0.05 were considered statistically significant.

**RESULTS**

**Search and inclusion**

Our initial search identified 245 studies, of which 35 met our inclusion criteria (see data supplement to the online version of this paper). However, 6 only met the inclusion criteria when additional information was provided. Furthermore, 2 unpublished studies (Hopwood et al; Kaloupek et al) were identified. In 3 studies we excluded the traumatised control group (Yehuda et al, 1995a, 1997), because several people in these groups had a history of PTSD.

A total of 1628 people were included across all studies, with 828 with PTSD and 800 controls. The median sample size of the studies was 20 for people with PTSD (range 7–75) and 18 for controls (range 7–113).

**Study characteristics**

Each study included in the meta-analysis and the assessed variables are shown in a data supplement to the online version of this paper. Cortisol was either assessed in plasma/serum (24 studies), saliva (8), or 24 h urine free cortisol samples (7), or a combination of two types of assessment.

In 7 studies, both trauma-exposed and non-exposed controls were used as comparison groups, whereas comparison solely with traumatised or non-traumatised controls was made in 9 and 8 studies respectively. In the remaining 13 studies trauma-exposed and non-exposed individuals were combined or left undefined in the comparison group. Studies reported on the following populations: combat veterans (18 studies), victims of (childhood) sexual or physical abuse (6), refugees (3), and various trauma (8). Two studies did not report information about the types of trauma in their population. Twenty-four studies matched their participants with PTSD and controls for gender and 8 studies matched them for age. Few studies matched their participants for other criteria such as smoking, race or menstrual cycle. Potential confounding factors such as medication usage, exclusion criteria and dietary restrictions varied greatly across studies (see data supplement to online version of this paper).

**Overall comparison**

Figure 1 shows a forest plot of the SMD of cortisol levels in people with PTSD relative to controls in each of the 37 studies, grouped according to the type of measurement. There was no overall difference in pooled effect size between people with PTSD and controls (SMD $= −0.12$, 95% CI $= −0.32$ to 0.08, $P = 0.24$). There was significant heterogeneity in results between studies: the $P$-value for the $Q$-test for heterogeneity beyond chance was 0.0001 and $I^2$ was 71% (indicating that 71% of variation across studies can be attributed to heterogeneity rather than chance). In the situation that all studies were measuring the same SMD, only sampling variation would be present and $I^2$ would be zero (Higgins et al, 2003).

**Type of assessment**

We analysed whether the SMD in cortisol levels between people with PTSD and controls depended on the type of measurement: plasma/serum (24 studies), saliva (7), and 24 h urinary free cortisol (6). None of the measurements revealed differences between patients and controls (plasma/serum, SMD $= −0.080$, 95% CI $= −0.32$ to 0.17; saliva, SMD $= −0.19$, 95% CI $= −0.63$ to 0.26; 24 h urine, SMD $= −0.20$, 95% CI $= −0.72$ to 0.31).

**Subgroup analyses**

All the following subgroup analyses were performed with studies measuring cortisol in plasma/serum. The small numbers of studies measuring cortisol in saliva and urine prohibited any meaningful subgroup analysis.

**Time of measurement**

In this analysis we grouped studies according to whether measurements were taken between 08.00 and 09.00 h or in the afternoon. No differences were found for morning samples ($n = 15$ studies, SMD $= −0.0006$, 95% CI $= −0.53$ to 0.53, $P = 0.998$), whereas in the afternoon people with PTSD had lower levels of cortisol than controls ($n = 7$ studies, SMD $= −0.79$, 95% CI $= −1.38$ to $−0.003$, $P = 0.049$). However, the formal test for interaction did not reach significance, so the observed effect in the afternoon subgroup should be interpreted with caution (Fig. 2).

**Gender**

The influence of gender on cortisol level was examined in 10 studies which included only males and in 4 studies with only females. Whereas the analyses of cortisol in males with and without PTSD did not reveal any significant differences (SMD $= 0.15$, 95% CI $= −0.06$ to 0.37, $P = 0.16$), females with PTSD had highly significant lower cortisol levels than their comparison groups (SMD $= −0.49$, 95% CI $= −0.86$ to $−0.13$, $P = 0.009$). Because of this large difference in effect, significant interaction between male and female studies was also found.
**Type of trauma**

Four categories were used to examine whether type of trauma might have influenced the association between PTSD and cortisol: war veterans, victims of sexual or physical abuse, refugees and various trauma (Fig. 2). Subgroup analysis revealed significantly lower cortisol levels in people with PTSD due to sexual or physical abuse than in controls (n=5 studies, SMD= -0.55, 95% CI -0.89 to -0.21, P=0.002). No differences in cortisol level were found between controls and people with PTSD due to other types of trauma (war veterans, n=12 studies, SMD= -0.15, 95% CI -0.03 to 0.34, P=0.12; refugees, n=2, SMD= -0.17, 95% CI -0.63 to 0.29, P=0.46; various trauma, n=3, SMD= -0.11, 95% CI -0.52 to 0.31, P=0.61). Analysis also revealed a significant interaction between types of trauma (n=22, F=4.31, d.f.=3, 1000, P=0.005).

**Years since trauma**

To analyse whether the number of years elapsed since the traumatic event was related to cortisol levels, we constructed three time categories: 0–10 years, 11–20 years and over 20 years. In studies where there were no such data, we approximated this time frame by using the duration of illness. For studies examining people with childhood abuse, we subtracted 12 years from their mean age. For those with war-related trauma we subtracted the year of study publication from the year in which that particular war ended. Among studies with the same time frame for years since trauma no differences were found for cortisol levels of people with PTSD and controls (0–10 years, n=8 studies, SMD=0.19, 95% CI -0.11 to 0.49, P=0.21; 11–20 years, n=2, SMD=0.52, 95% CI -0.058 to 1.09, P=0.078; >20 years, n=9, SMD=0.19, 95% CI -0.44 to 0.071, P=0.16).

**Exposure to trauma of control groups**

To differentiate between exposure to trauma and exposure with subsequent development of PTSD, we analysed studies which indicated whether controls had previous exposure to trauma. In 17 studies we calculated effect sizes for PTSD compared with trauma-exposed controls and non-exposed controls separately. Lower cortisol levels were found for people with PTSD compared with non-exposed controls (n=11 studies, SMD= -0.35, 95% CI -0.61 to -0.098, P=0.007). No differences were found between people with PTSD and trauma-exposed controls (n=9, SMD= 0.096, 95% CI -0.16 to 0.35, P=0.46). The test of interaction was also significant (n=20, F=5.93, d.f.=1, 1000, P=0.015).

**Comorbid depression**

The effect of comorbid depression in people with PTSD on cortisol level was analysed in 13 studies that reported whether depression was present or absent within their PTSD group. If a single study used two subgroups with PTSD, with and without comorbid depression, we included both comparisons in the analysis (2 studies). The results (see Fig. 2) showed that depression had no influence on the effect sizes.

**Year of publication**

Year of publication was examined to evaluate whether improved methodology might strengthen possible contrasts between groups. Using all published studies, year of publication was linearly modelled to estimate the change in SMD per year. Figure 2 shows the effect on SMD over a 10-year period (1994–2004). Results could not confirm this hypothesis.

## DISCUSSION

**Main findings**

In this systematic review we included 37 studies examining basal cortisol levels in adults with current PTSD in comparison with adults without psychiatric disorders published between 1980 and March 2005. Combining all available data for meta-analysis we found no systematic difference in basal cortisol levels between people with PTSD and controls. However, results were highly heterogeneous, indicating that differences between subgroups might be present. Subsequent explanatory subgroup analyses revealed that studies assessing plasma or serum reported significantly lower cortisol levels in people with PTSD compared with controls when compared with controls with no previous exposure to trauma. Lower cortisol levels were also found in people with PTSD compared with controls in studies including only females, in studies on physical or sexual abuse and in afternoon
Influence of variables on plasma/serum cortisol assessments. Standardised mean difference (with 95% CI) of plasma/serum cortisol levels between people with post-traumatic stress disorder (PTSD) and controls.

<table>
<thead>
<tr>
<th>Type of assessment</th>
<th>Control higher</th>
<th>PTSD higher</th>
<th>P-value for interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of measurement</td>
<td>Morning (n=15)</td>
<td>Afternoon (n=7)</td>
<td>0.102</td>
</tr>
<tr>
<td>Gender</td>
<td>Female (n=4)</td>
<td>Male (n=10)</td>
<td>0.003</td>
</tr>
<tr>
<td>Trauma type</td>
<td>War veterans (n=12)</td>
<td>Sexual/physical abuse (n=5)</td>
<td>Refugees (n=2)</td>
</tr>
<tr>
<td>Years since trauma</td>
<td>0–10 (n=6)</td>
<td>11–20 (n=2)</td>
<td>&gt;20 (n=9)</td>
</tr>
<tr>
<td>Concomitant depression</td>
<td>Depression absent (n=9)</td>
<td>Depression present (n=4)</td>
<td>0.418</td>
</tr>
<tr>
<td>Exposure controls</td>
<td>Not exposed (n=11)</td>
<td>Trauma exposed (n=9)</td>
<td>0.015</td>
</tr>
<tr>
<td>Publication year</td>
<td>1994</td>
<td>2004</td>
<td>0.983</td>
</tr>
</tbody>
</table>

Despite possible abnormal CBG levels in studies using plasma/serum measurements, differences in study results did not depend on body fluid used for cortisol assessment.

### Time of assessment

Since cortisol has a circadian rhythm, with low values at awakening, followed by peak values 30 min after awakening and a steady decline during the rest of the day, time of assessment is expected to be an important factor in its measurement. Our findings revealed that during the afternoon, people with PTSD had significantly lower cortisol levels than controls. No such differences were found during the early morning. Since PTSD is known to be associated with difficulties in sleeping, and measures depended on fixed times, it is uncertain whether variability in awakening time between PTSD and controls influenced morning cortisol levels. Since cortisol secretion is relatively stable during the afternoon actual differences can be detected more easily.

### Trauma-exposed v. non-exposed control groups

Some of the disparity in study results can be explained by whether studies used trauma-exposed or non-exposed controls. We found significantly lower plasma/serum cortisol levels in people with PTSD compared with controls not exposed to trauma but no such difference when comparisons were made with trauma-exposed controls. This suggests that differences in cortisol levels relate to being exposed to trauma generally rather than to PTSD.

### Gender

In general, women appear to have a more sensitised HPA axis with lower overall plasma cortisol than men (Van Cauter et al, 1996), and our findings indicated that females with PTSD showed lower levels of basal cortisol than female controls. No such difference was apparent between males. This may explain why women are more vulnerable than men to the development of post-trauma symptoms and take longer to recover from them. In addition, the higher risk of PTSD in women may be due, at least in part, to the types of traumas they experience (more interpersonal violence, particularly of a sexual nature), to higher peri-traumatic dissociation in women, or to women’s use of avoidant coping strategies (Olff et al, 2007). Gender-specific...
PTSD subgroups may exist – in particular arousal-related and dissociation-related variants – with distinct neuropsychological profiles and attendant symptoms. Hence, gender-specific psychobiological reactions to trauma may contribute to the higher risk for PTSD.

**Physical or sexual abuse and years since trauma**

Only people with PTSD due to physical or sexual abuse had lower cortisol levels than controls. This type of trauma is generally chronic and often starts in early development. An upbringing that is associated with adversity can produce detrimental effects on health (Fish et al., 2004) and it is likely that within a critical phase during development, functioning of the HPA axis alters. Years elapsed since trauma had no pronounced influence on cortisol levels of people with PTSD. The time of onset of PTSD in development and the ongoing traumatising character of abuse might be more crucial in distinguishing abuse-related PTSD from other types of trauma. It should be noted that for statistical analyses we could not disentangle female gender from victims of abuse because of overlap in studies. Therefore, we could not examine whether low cortisol levels found in women with PTSD are due to gender or type of trauma preceding PTSD, or an interaction of both.

**Comorbid depression**

Comorbid depression in people with PTSD had no influence on the association between PTSD and cortisol level. Although there seems to be consensus that people with depression demonstrate high cortisol levels (Holsboer, 2001), generally this hyperactivation of the HPA axis is typically found in severe depression not specifically due to traumatic stress. In accordance with our findings, PTSD and comorbid PTSD/depression following traumatic injury were indistinguishable and reflected a shared vulnerability with a range of similar predictive non-biological variables. Comorbid PTSD/depression and PTSD alone may reflect one and the same construct (O’Donnell et al., 2004), as appears to be confirmed by similar cortisol values within the present study.

**Year of publication/sensitivity of assays**

We also examined whether year of publication affected the relationship between PTSD and cortisol. This variable was included in our subgroup analysis to serve as a proxy for changes in protocol or the use of more sensitive assays during the study period. Year of publication had no impact on the results and therefore did not explain any of the heterogeneity in results across studies.

**Limitations and future research**

We acknowledge several limitations of our meta-analyses. First, despite our efforts to include unpublished studies, publication bias might still have obscured our results, as studies which find significant differences are more likely to be published. Second, we performed several subgroup analyses based on characteristics of the PTSD or control group and used these as study-level covariates in our model. Owing to incomplete reporting, we could not use all studies in the subgroup analyses. In some cases, only a few studies were available within a specific stratum, which caused problems related to chance findings and lack of power. Subgroup analyses within systematic reviews have to be interpreted with care because by nature they are post hoc analyses. They can provide additional insight but have to be confirmed in well-designed prospective studies with sufficient power to examine differences in effect by subgroups. Furthermore, individual appraisal and coping mechanisms are crucial in determining levels of stress hormones such as cortisol (Olff et al., 2005a,b), factors that have not been systematically assessed in most of the literature. Finally, substantial differences in the methodology of studies hampered comparison. For instance, restrictions regarding smoking, alcohol, drugs and medication usage varied widely across studies. Although we attempted to pool data of studies on important characteristics, these restrictions and comorbid conditions known to confound cortisol levels remained unattended.

In future studies consensus in data collection and sampling protocol of basal diurnal cortisol would facilitate comparison of data across studies. Given that saliva samples can be obtained by study participants themselves in their own environment and related to time of awakening and that salivary cortisol consists completely of the bioactive fraction, future studies could overcome several difficulties by sampling salivary cortisol.

In summary, across 37 studies people with PTSD and healthy controls did not differ in cortisol levels. Nevertheless, support was found for low cortisol levels dependent on the type of control group and specific sub-populations. Significantly lower cortisol levels were found in people with PTSD when compared with non-exposed controls, whereas no such differences were found when compared with trauma-exposed controls without PTSD. Subgroup analyses further revealed lower cortisol in people who seem to be at the greatest risk for developing PTSD, i.e. women and physically or sexually abused victims. The lower cortisol values in PTSD found in the afternoon endorse the need to choose the time of measurement carefully. It is important to note that numerous factors – which are frequently overlooked – may have a confounding influence on cortisol levels. Therefore, disentangling the relationship between PTSD and cortisol is more complex than it first appears.

**REFERENCES**


American Psychiatric Association (1980) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn) (DSM—III). APA.


Neuropsychobiology (2003) 327


MARI LOUISE MEEWISSE, MSc, Center for Psychological Trauma, Department of Psychiatry and Academic Medical Center de Meeren, University of Amsterdam; JOHANNE S B. REITSM A, PhD, Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam; GIEL JAN DE VRIES, MSc, BERTHOLD P. R. GERSONS, MD, PhD; MIRANDA OLF, PhD, Center for Psychological Trauma, Department of Psychiatry and Academic Medical Center de Meeren, University of Amsterdam, The Netherlands

Correspondence: Marie-Louise Meewisse, Center for Psychological Trauma, Academic Medical Center de Meeren, Department of Psychiatry, University of Amsterdam, Meibergdreef 5, 1105 AZ Amsterdam, The Netherlands. Email: M.L.Meeewisse@azc.uva.nl

(First received 31 March 2006, final revision 17 March 2007, accepted 2 May 2007)
Prevalence of depressive symptoms and syndromes in later life in ten European countries

The SHARE study


Background The EURO–D, a 12-item self-report questionnaire for depression, was developed with the aim of facilitating cross-cultural research into late-life depression in Europe.

Aims To describe the national variation in depression symptoms and syndrome prevalence across ten European countries.

Method The EURO–D was administered to cross-sectional nationally representative samples of non-institutionalised persons aged ≥ 50 years (n=22,777). The effects of age, gender, education and cognitive functioning on individual symptoms and EURO–D factor scores were estimated. Country-specific depression prevalence rates and mean factor scores were re-estimated, adjusted for these compositional effects.

Results The prevalence of all symptoms was higher in the Latin ethno-lingual group of countries, especially symptoms related to motivation. Women scored higher on affective suffering; older people and those with impaired verbal fluency scored higher on motivation.

Conclusions The prevalence of individual EURO–D symptoms and of probable depression (cut-off score ≥ 4) varied consistently between countries. Standardising for effects of age, gender, education and cognitive function suggested that these compositional factors did not account for the observed variation.

Declaration of interest None. Funding detailed in Acknowledgements.

Depression is common in later life. However, there is considerable variation in reported prevalence between studies worldwide. There have been relatively few direct cross-national comparisons of the prevalence of depression using comparable methodology, particularly with respect to sampling, definition and assessment of outcome. Methodological differences between studies preclude firm conclusions about cross-cultural and geographical variation (Beekman et al, 1999). Improving the comparability of epidemiological research constitutes an important step forward. The EURO–D scale (Prince et al, 1999b) was developed to harmonise data on late-life depression from 11 European population-based studies (EURODEP). The EURO–D scale has more recently been administered in a large, collaborative household survey of nationally representative samples of people aged 50 years and over from ten European countries: the Survey of Health, Ageing and Retirement in Europe (SHARE). Observed differences in prevalence of depression in later life may be accounted for by methodological, compositional or contextual factors. In this analysis we sought to answer three questions:

(a) Are there differences in the prevalence of depression between European countries, and are these consistent across the two factors that underlie the EURO–D measure and its 12 constituent symptom-based items?

(b) Are there compositional differences between the older European populations in terms of age and gender (found previously to be associated with the ‘motivation’ and ‘affective suffering’ factors respectively) (Prince et al, 1999a), education, and cognitive function (previously hypothesised to be associated with motivation factor) (Prince et al, 1999a)?

(c) Do these compositional differences account, wholly or partly, for any observed differences in depression prevalence and EURO–D scale scores?

METHOD

Survey design

The SHARE study (Borsch-Supan et al, 2005) is a consortium survey of health in older people across Europe. In this study, national survey organisations were responsible for selecting household samples and conducting interviews in nationally representative samples of people aged 50 years and over from ten countries: Denmark, Sweden, The Netherlands, Germany, Austria, Switzerland, France, Spain, Italy and Greece.

The SHARE interview was specifically designed to cross-link with the US Health and Retirement Study (http://hrsonline.isr.umich.edu) and the English Longitudinal Study of Ageing (http://www.elan.ac.uk/elsa), with the advantage that it encompasses international variation in culture, health and social welfare systems and public policy. Questions covered health variables (self-reported health, physical functioning, cognitive functioning, health behaviour, use of healthcare facilities), psychological variables (depression, wellbeing, life satisfaction), economic variables (current work activity, job characteristics, opportunities to work past retirement age, sources and composition of current income, wealth and consumption, housing and education) and social support variables (assistance within families, transfers of income and assets, social networks, volunteer activities). All of the above topics were rated in an interview conducted in the respondent’s home, with an average interview duration of around 90 min. Response rates were acceptable throughout. The data are freely available to the research community (http://www.share-project.org).

Measurements

The EURO–D was originally developed to compare symptoms of depression in 11 European centres (Prince et al, 1999b). Its items are derived from the Geriatric Mental State examination (GMS; Copeland et al, 1986) and cover 12 symptom domains: depressed mood, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. Each item is scored 0 (symptom not present) or 1 (symptom present), and item scores are summed to produce a scale with
a minimum score of zero and a maximum of 12.

The psychometric properties of the EURO-D have been extensively investigated and criterion validity demonstrated in the cross-cultural context. Principal components analysis generated two factors (affective suffering and motivation) that were common to nearly every participating European country in the EURODEP studies (Prince et al, 1999b) and for Indian, Latin-American and Caribbean centres in the 10/66 Dementia Research Group pilot studies (Prince et al, 2004). Subsequent analysis of the EURO-D in the SHARE data-set using confirmatory factor analysis confirmed the two-factor solution of the EURO-D and suggested measurement invariance across the ten countries (common factor loadings and item calibrations), at least for the ‘affective suffering’ factor (Castro-Costa et al, 2007). Criterion validity for this measure was demonstrated in each of the EURODEP study sites, with an optimal cut-off point of a score of 4 or above against a variety of criteria for clinically significant depression (Prince et al, 1999b). The EURO-D was also found to be reliable and was validated against the criterion of DSM–III–R depression in older people in Spain (Larraga et al, 2006).

The following aspects of cognitive function were measured in all participants: memory, using delayed recall of a ten-word list in wide international use (Ganguli et al, 1996; Prince et al, 2003), the only difference being that in our study this was presented once only in the learning phase, as opposed to the conventional three presentations; and verbal fluency, measured using the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) animal naming task (Goodglass & Kaplan, 1983). Other factors considered in the analysis were age, gender and duration of education.

Statistical analyses
Country-specific prevalences of all 12 EURO-D items were derived, as were prevalence of EURO-D scores of 4 or more, and country-specific mean scores for the affective suffering and motivation sub-scales. The independent effects of gender, age (<75 years v. 75+ years), duration of education (11+ years v. <11 years), verbal fluency score (<10 v. 10+ animals named) and memory (3+ v. <3 words recalled) upon individual symptoms were estimated in each country as mutually adjusted prevalence ratios from Poisson regression models with 95% confidence intervals. For each covariate, an effect by country interaction term was added to the final model to test for heterogeneity. Associations with factor scores for the two sub-scales were estimated as eta-squared statistics derived from generalised linear modelling (GLM), again with effect by country interaction terms fitted in the final stage. Finally, ‘affective suffering’ and ‘motivation’ scores were estimated, adjusting for age, gender, education, verbal fluency and memory using GLM, and the country-specific prevalence of EURO-D depression (a score of 4 or over) was standardised separately for age, gender, education, verbal fluency and memory and for all effects simultaneously, using the direct method to the age, gender, education and verbal fluency distribution of the pooled data-set. All analyses were conducted with Stata version 9.1 for Windows.

RESULTS

Based on probability samples in all participating countries, the SHARE sampling strategy aimed to represent the non-institutionalised population aged 50 years and older. The SHARE investigators have compared sample characteristics with other data sources, indicating that although there were some expected divergences, this objective has been broadly achieved (Borsch-Supan et al, 2005).

The characteristics of the sample by country, gender and age and also household and individual response rates have been reported in detail elsewhere (Borsch-Supan et al, 2005). The proportion of households responding was 57.4% overall, with the lowest response in Switzerland (37.6%) and the highest in France (69.4%). Individual response proportions ranged from 73.8% (Italy) to 93.0% (Denmark), with a rate of 86.0% overall. The principal characteristics of the respondents in each country are summarised in Table 1. The distribution of gender and age did not differ between countries. In the pooled data-set, 54.5% were female, and the mean age was 64.7 years (s.d.=10.0). Educational levels were lowest in the Latin countries (France, Italy and Spain) and in Greece. Participants from these countries also recorded the lowest (most impaired) scores on both the animal naming task and the delayed recall of the ten-word delayed recall list. In most countries more than half of the sample were retired, the exceptions being The Netherlands (32%), Spain (35%), Switzerland (45%) and Greece (45%). Mean EURO-D scores were statistically different between countries (F=68.79; P<0.00001) and were highest in France, Spain and Italy. The prevalence rates of individual depressive symptoms are displayed in Figs 1 and 2 for symptoms loading principally on affective suffering and motivation respectively. These varied consistently and significantly between countries (P<0.001) for all 12 symptoms, with a higher prevalence in France, Spain and Italy. Among affective suffering symptoms, depressed mood, tearfulness, fatigue and sleep disturbance were most common. Among motivation symptoms, poor concentration was reported most frequently.

Associations with affective suffering symptoms are summarised in Table 2: depression, tearfulness, suicidality and fatigue were selected because they have the highest factor loading for affective suffering (Castro-Costa et al, 2007). The four individual symptoms and the overall factor score were each consistently associated with gender, with higher prevalence of symptoms and higher factor scores among women, with negligible influence of age, education or cognitive function. None of these effects varied significantly between countries.

Associations with motivation symptoms are summarised in Table 3. enjoyment, pessimism and interest were chosen because of their high factor loading for motivation (Castro-Costa et al, 2007). The three individual symptoms and the overall factor score were consistently and strongly associated with age, with a higher prevalence of symptoms and higher factor scores among older people. Effects of gender and education were negligible. Motivation symptoms and factor score were also strongly associated with lower verbal fluency, but were not associated with impaired memory. None of these associations varied significantly between countries.

The prevalence of case-level depression according to the EURO-D scale is summarised and compared between nations in Table 4. Consistent with the observations for individual EURO-D symptoms, the highest prevalence rates were found in France, Italy and Spain, with a 9% difference between the lowest of these (33% in France) and the next lowest (24% in Greece). Prevalence in the remaining countries was...
## Table 1  Demographic and cognitive variables of the sample (n = 22,777)

<table>
<thead>
<tr>
<th></th>
<th>Sweden</th>
<th>Denmark</th>
<th>Netherlands</th>
<th>Germany</th>
<th>Austria</th>
<th>Switzerland</th>
<th>France</th>
<th>Italy</th>
<th>Spain</th>
<th>Greece</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 3067)</td>
<td>(n = 1732)</td>
<td>(n = 3000)</td>
<td>(n = 1986)</td>
<td>(n = 1010)</td>
<td>(n = 1842)</td>
<td>(n = 2559)</td>
<td>(n = 2419)</td>
<td>(n = 2142)</td>
<td></td>
</tr>
<tr>
<td><strong>Age, years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age group, %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59</td>
<td>35.4</td>
<td>39.9</td>
<td>41.7</td>
<td>34.0</td>
<td>30.6</td>
<td>37.0</td>
<td>40.3</td>
<td>33.0</td>
<td>31.1</td>
<td>37.5</td>
</tr>
<tr>
<td>60–74</td>
<td>44.8</td>
<td>39.5</td>
<td>42.3</td>
<td>49.5</td>
<td>50.6</td>
<td>41.7</td>
<td>38.7</td>
<td>51.7</td>
<td>44.5</td>
<td>42.6</td>
</tr>
<tr>
<td>≥ 75</td>
<td>19.7</td>
<td>20.5</td>
<td>16.0</td>
<td>16.5</td>
<td>18.8</td>
<td>21.2</td>
<td>21.0</td>
<td>15.3</td>
<td>24.4</td>
<td>19.8</td>
</tr>
<tr>
<td><strong>Mean (range)</strong></td>
<td>65 (50–102)</td>
<td>64 (50–104)</td>
<td>63 (50–97)</td>
<td>65 (50–100)</td>
<td>65 (50–96)</td>
<td>64 (50–99)</td>
<td>64 (50–100)</td>
<td>66 (50–103)</td>
<td>64 (50–97)</td>
<td></td>
</tr>
<tr>
<td><strong>Female, %</strong></td>
<td>53.57</td>
<td>54.68</td>
<td>54.11</td>
<td>58.71</td>
<td>53.66</td>
<td>56.89</td>
<td>55.76</td>
<td>58.50</td>
<td>57.94</td>
<td></td>
</tr>
<tr>
<td><strong>Married, %</strong></td>
<td>68.78</td>
<td>62.68</td>
<td>77.28</td>
<td>75.21</td>
<td>66.89</td>
<td>77.00</td>
<td>72.26</td>
<td>69.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education, years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>10.37</td>
<td>12.71</td>
<td>11.11</td>
<td>13.53</td>
<td>11.33</td>
<td>12.17</td>
<td>8.20</td>
<td>7.03</td>
<td>5.48</td>
<td>8.67</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>10</td>
<td>14</td>
<td>10</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>5.0</td>
<td>4.5</td>
<td>6</td>
</tr>
<tr>
<td><strong>Education, years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>25%</strong></td>
<td>7</td>
<td>10.25</td>
<td>10</td>
<td>13</td>
<td>8</td>
<td>12</td>
<td>5</td>
<td>5.0</td>
<td>4.5</td>
<td>6</td>
</tr>
<tr>
<td><strong>75%</strong></td>
<td>13</td>
<td>15</td>
<td>14</td>
<td>16.5</td>
<td>12</td>
<td>16</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td><strong>Retired, %</strong></td>
<td>52.61</td>
<td>50.85</td>
<td>32.52</td>
<td>51.10</td>
<td>64.50</td>
<td>45.60</td>
<td>50.74</td>
<td>54.51</td>
<td>35.04</td>
<td>45.45</td>
</tr>
<tr>
<td><strong>Euro–D score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean (s.d.)</strong></td>
<td>1.9 (1.9)</td>
<td>1.8 (1.9)</td>
<td>1.9 (2.0)</td>
<td>1.9 (2.0)</td>
<td>1.9 (2.1)</td>
<td>1.9 (1.8)</td>
<td>2.8 (2.3)</td>
<td>2.8 (2.5)</td>
<td>3.1 (2.8)</td>
<td>2.1 (2.2)</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>25%</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>75%</strong></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Affective suffering: mean (s.d.)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20 (0.33)</td>
<td>0.27 (0.30)</td>
<td>0.24 (0.32)</td>
<td>0.29 (0.36)</td>
<td>0.28 (0.33)</td>
<td>0.27 (0.35)</td>
<td>0.30 (0.35)</td>
<td>0.27 (0.33)</td>
<td>0.24 (0.32)</td>
<td>0.26 (0.32)</td>
<td></td>
</tr>
<tr>
<td><strong>Motivation: mean (s.d.)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>− 0.07 (0.27)</td>
<td>0.07 (0.27)</td>
<td>0.11 (0.28)</td>
<td>0.03 (0.30)</td>
<td>0.01 (0.32)</td>
<td>0.18 (0.31)</td>
<td>0.07 (0.29)</td>
<td>0.09 (0.28)</td>
<td>0.11 (0.28)</td>
<td>0.09 (0.29)</td>
<td></td>
</tr>
<tr>
<td><strong>Animal naming &lt;10, %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.8</td>
<td>19.5</td>
<td>26.0</td>
<td>25.9</td>
<td>30.5</td>
<td>23.6</td>
<td>37.5</td>
<td>50.2</td>
<td>57.1</td>
<td>32.5</td>
<td></td>
</tr>
<tr>
<td><strong>Ten-word list &lt;3, %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>4.4</td>
<td>3.7</td>
<td>4.9</td>
<td>5.1</td>
<td>2.8</td>
<td>7.8</td>
<td>23.1</td>
<td>17.7</td>
<td>13.6</td>
<td></td>
</tr>
</tbody>
</table>
Prevalence of affective suffering symptoms. Differences were not affected by direct standardisation for gender, age, education, verbal fluency or memory.

DISCUSSION

In this study the prevalence of individual EURO–D symptoms varied consistently between countries, with a higher prevalence in three Latin countries – France, Italy and Spain. Prevalence of probable depression according to the EURO–D cut-off score of 4 or more therefore followed the same distribution. The distribution of age and gender was similar across the ten countries, but educational levels were lower and cognitive function more impaired in France, Italy, Spain and Greece. Standardising for the effects of age, gender, education and cognitive function suggested that compositional differences in these factors did not account for the observed variation in the prevalence of depression.

Strengths and weaknesses of the study design

This study has the advantage of using data from nationally representative samples of those aged 50 years and over from ten European countries. Strong conceptual validity, high internal consistency and a common factor structure across different European centres was previously demonstrated for the depression assessment, the EURO–D (Prince et al., 1999b). Furthermore, these favourable psychometric properties were confirmed in an earlier analysis of SHARE data using more advanced psychometric techniques – confirmatory factor analysis and Rasch modelling – to support the cross-cultural validity of the measure (Castro-Costa et al., 2007). These analyses provided further robust evidence to support a two-factor solution: affective suffering (well characterised, and invariant across cultures) and motivation (less well characterised and variable across cultures). As with the mental health survey undertaken by the European Study of the Epidemiology of Mental Disorders (ESEMed; Alonso et al., 2004), the SHARE data are limited by the relatively low proportion of households and individuals responding. This may, unfortunately, represent a secular trend in more developed countries. The net effect may be an underestimation of the true prevalence of depression (Eaton et al., 1992; De Graaf et al., 2000). We used a simple scale-based assessment for depression rather than a comprehensive clinical diagnostic interview. Nevertheless, the EURO–D and its cut-off point of 4 or more have previously been validated against relevant clinical assessments in different European settings (Prince et al., 1999b).

Consistency with findings from other research

Findings from the SHARE survey are most directly comparable with those of the EURODEP consortium studies, in which the same outcome measure (the EURO–D) was administered to older adults in cross-sectional population-based surveys. In descending order, the mean EURO–D scores for each of the EURODEP centres that used the GMS were: Munich (Germany), 3.58; London (UK), 2.54; Berlin (Germany), 2.48; Iceland, 2.03; Amsterdam (The Netherlands), 1.98; Verona (Italy), 1.84; Liverpool (UK), 1.79; Zaragoza (Spain), 1.61; and Dublin (Ireland), 1.34 (Prince et al., 1999a). The EURODEP findings do not, therefore, support our finding of higher levels of reported depression symptoms in Latin countries. However, there are important differences between the SHARE and EURODEP studies. First, none of the EURODEP centres used nationally representative samples, and the age range was 65 years and over rather than 50 years and over as in SHARE. Second, the EURO–D items were nested within the more comprehensive GMS clinical interview. Finally, in several centres the GMS was administered by clinicians working for university research groups rather than by lay interviewers working for survey organisations as with SHARE. Nevertheless, several findings in our analysis are consistent with those of EURODEP: motivation factor scores and EURO–D scores increase with age, and affective suffering scores and EURO–D scores are higher in women than in men (Prince et al., 1999a). The EURODEP investigators postulated that some of the effect of age on motivation factor scores might have been accounted for by...
<table>
<thead>
<tr>
<th></th>
<th>Sweden</th>
<th>Denmark</th>
<th>Netherlands</th>
<th>Germany</th>
<th>Austria</th>
<th>Switzerland</th>
<th>France</th>
<th>Italy</th>
<th>Spain</th>
<th>Greece</th>
<th>Test for interaction (effect × country), P</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (F v. M)</td>
<td>1.02</td>
<td>1.09</td>
<td>1.46</td>
<td>1.34</td>
<td>1.95</td>
<td>1.60</td>
<td>1.35</td>
<td>1.51</td>
<td>1.61</td>
<td>1.01</td>
<td>0.93 (0.62-1.42)</td>
<td></td>
</tr>
<tr>
<td>Age (≥75 v. &lt;75)</td>
<td>1.05</td>
<td>0.87</td>
<td>1.07</td>
<td>0.84</td>
<td>1.05</td>
<td>0.96</td>
<td>0.76</td>
<td>1.10</td>
<td>0.82</td>
<td>1.10</td>
<td>0.96 (0.68-1.37)</td>
<td></td>
</tr>
<tr>
<td>Education (&gt;11 v. 11 years)</td>
<td>1.23</td>
<td>1.08</td>
<td>1.03</td>
<td>0.83</td>
<td>0.97</td>
<td>0.87</td>
<td>0.72</td>
<td>0.94</td>
<td>1.06</td>
<td>0.83</td>
<td>0.81 (0.65-1.00)</td>
<td></td>
</tr>
<tr>
<td>Animal naming (&lt;10 v. ≥10)</td>
<td>1.05</td>
<td>0.69</td>
<td>1.00</td>
<td>0.65</td>
<td>0.91</td>
<td>0.88</td>
<td>0.76</td>
<td>0.93</td>
<td>1.04</td>
<td>1.00</td>
<td>0.92 (0.78-1.10)</td>
<td></td>
</tr>
<tr>
<td>Ten-word list (&gt;3 v. &lt;3)</td>
<td>0.86</td>
<td>0.73</td>
<td>0.89</td>
<td>0.70</td>
<td>0.97</td>
<td>0.87</td>
<td>0.76</td>
<td>0.79</td>
<td>0.85</td>
<td>0.96</td>
<td>0.89 (0.74-1.07)</td>
<td></td>
</tr>
<tr>
<td><strong>Suicidality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (F v. M)</td>
<td>0.90</td>
<td>0.71</td>
<td>0.76</td>
<td>0.53</td>
<td>1.01</td>
<td>1.00</td>
<td>0.76</td>
<td>0.85</td>
<td>1.00</td>
<td>1.15</td>
<td>1.10 (0.92-1.33)</td>
<td></td>
</tr>
<tr>
<td>Age (≥75 v. &lt;75)</td>
<td>0.90</td>
<td>0.71</td>
<td>0.76</td>
<td>0.53</td>
<td>1.01</td>
<td>1.00</td>
<td>0.76</td>
<td>0.85</td>
<td>1.00</td>
<td>1.15</td>
<td>1.10 (0.92-1.33)</td>
<td></td>
</tr>
<tr>
<td>Education (&gt;11 v. 11 years)</td>
<td>1.20</td>
<td>1.04</td>
<td>0.90</td>
<td>0.68</td>
<td>0.86</td>
<td>0.93</td>
<td>0.80</td>
<td>0.79</td>
<td>1.01</td>
<td>1.00</td>
<td>1.00 (0.90-1.10)</td>
<td></td>
</tr>
<tr>
<td>Animal naming (&lt;10 v. ≥10)</td>
<td>0.96</td>
<td>0.57</td>
<td>1.32</td>
<td>0.67</td>
<td>1.13</td>
<td>0.86</td>
<td>1.07</td>
<td>0.89</td>
<td>1.04</td>
<td>1.01</td>
<td>1.01 (0.90-1.10)</td>
<td></td>
</tr>
<tr>
<td>Ten-word list (&gt;3 v. &lt;3)</td>
<td>0.94</td>
<td>0.68</td>
<td>0.88</td>
<td>0.63</td>
<td>1.02</td>
<td>0.87</td>
<td>0.84</td>
<td>0.89</td>
<td>1.03</td>
<td>0.82</td>
<td>0.88 (0.68-1.08)</td>
<td></td>
</tr>
<tr>
<td><strong>Fatigue</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (F v. M)</td>
<td>1.26</td>
<td>1.27</td>
<td>1.40</td>
<td>0.89</td>
<td>1.51</td>
<td>1.05</td>
<td>1.10</td>
<td>0.43</td>
<td>1.09</td>
<td>0.98</td>
<td>0.93 (0.69-1.38)</td>
<td></td>
</tr>
<tr>
<td>Age (≥75 v. &lt;75)</td>
<td>1.26</td>
<td>1.06</td>
<td>1.22</td>
<td>0.94</td>
<td>1.35</td>
<td>1.15</td>
<td>1.05</td>
<td>1.00</td>
<td>1.15</td>
<td>0.98</td>
<td>0.93 (0.69-1.38)</td>
<td></td>
</tr>
<tr>
<td>Education (&gt;11 v. 11 years)</td>
<td>0.97</td>
<td>0.85</td>
<td>0.94</td>
<td>0.87</td>
<td>1.10</td>
<td>0.90</td>
<td>0.79</td>
<td>1.15</td>
<td>1.00</td>
<td>0.96</td>
<td>0.91 (0.69-1.38)</td>
<td></td>
</tr>
<tr>
<td>Animal naming (&lt;10 v. ≥10)</td>
<td>0.96</td>
<td>0.62</td>
<td>0.81</td>
<td>0.55</td>
<td>0.70</td>
<td>0.91</td>
<td>0.62</td>
<td>0.80</td>
<td>0.97</td>
<td>0.93</td>
<td>0.96 (0.69-1.38)</td>
<td></td>
</tr>
<tr>
<td>Ten-word list (&gt;3 v. &lt;3)</td>
<td>0.90</td>
<td>0.66</td>
<td>0.87</td>
<td>0.72</td>
<td>0.84</td>
<td>0.80</td>
<td>0.77</td>
<td>0.77</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00 (0.79-1.28)</td>
<td></td>
</tr>
</tbody>
</table>

F: female, M: male.
1. Prevalence ratio adjusted for other effects using Poisson regression.
2. Adjusted for other effects using generalised linear modelling.
### Table 3  Effects of gender, age, education, verbal fluency and memory upon items loading on motivation factor, by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Enjoyment</th>
<th>Pessimism</th>
<th>Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender (F v. M)</td>
<td>Age (≥75 v. &lt;75)</td>
<td>Education (≥11 v. &lt;11 years)</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.78 (0.61–0.98)</td>
<td>1.38 (1.03–1.86)</td>
<td>0.39 (0.25–0.61)</td>
</tr>
<tr>
<td>Denmark</td>
<td>1.00 (0.72–1.39)</td>
<td>1.18 (0.79–1.76)</td>
<td>0.58 (0.32–1.04)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1.04 (0.83–1.29)</td>
<td>1.34 (1.01–1.78)</td>
<td>0.53 (0.35–0.80)</td>
</tr>
<tr>
<td>Germany</td>
<td>1.06 (0.82–1.37)</td>
<td>0.85 (0.59–1.22)</td>
<td>0.66 (0.45–0.96)</td>
</tr>
<tr>
<td>Austria</td>
<td>0.78 (0.51–1.18)</td>
<td>1.55 (1.19–2.04)</td>
<td>0.88 (0.69–1.13)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.23 (1.05–1.43)</td>
<td>0.95 (0.55–1.64)</td>
<td>0.88 (0.54–1.43)</td>
</tr>
<tr>
<td>France</td>
<td>1.06 (0.79–1.43)</td>
<td>1.53 (1.08–2.15)</td>
<td>1.05 (0.76–1.43)</td>
</tr>
<tr>
<td>Italy</td>
<td>1.26 (0.99–1.60)</td>
<td>1.33 (1.09–1.63)</td>
<td>0.76 (0.51–0.96)</td>
</tr>
<tr>
<td>Spain</td>
<td>1.26 (0.99–1.60)</td>
<td>1.68 (1.34–2.12)</td>
<td>0.56 (0.36–0.87)</td>
</tr>
<tr>
<td>Greece</td>
<td>0.87 (0.66–1.12)</td>
<td>0.87 (0.65–1.19)</td>
<td>0.99 (0.83–1.17)</td>
</tr>
</tbody>
</table>

Effect size for association with EURO-D symptom. Prevalence ratio (95% CI).  
1. Adjusted for other effects using generalised linear modelling.
2. Adjusted for other effects using Poisson regression.
<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Education level</th>
<th>Verbal fluency</th>
<th>Animal naming</th>
<th>Standardised for the 5 effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>1.92 (1.78–2.06)</td>
<td>2.01 (1.93–2.24)</td>
<td>2.22 (2.02–2.42)</td>
<td>1.84 (1.68–1.99)</td>
<td>2.56 (2.16–2.96)</td>
</tr>
<tr>
<td>Denmark</td>
<td>1.81 (1.62–1.99)</td>
<td>1.91 (1.71–2.10)</td>
<td>1.93 (1.68–1.98)</td>
<td>1.86 (1.65–2.06)</td>
<td>2.32 (1.93–2.71)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1.94 (1.80–2.08)</td>
<td>2.06 (1.90–2.21)</td>
<td>2.19 (1.97–2.41)</td>
<td>1.88 (1.73–2.02)</td>
<td>2.44 (2.12–2.76)</td>
</tr>
<tr>
<td>Germany</td>
<td>1.85 (1.71–1.99)</td>
<td>2.01 (1.86–2.69)</td>
<td>2.34 (2.12–2.56)</td>
<td>2.72 (2.34–3.10)</td>
<td>2.61 (2.32–2.90)</td>
</tr>
<tr>
<td>Austria</td>
<td>1.93 (1.78–2.13)</td>
<td>2.06 (1.88–2.25)</td>
<td>2.36 (2.11–2.62)</td>
<td>1.75 (1.55–1.94)</td>
<td>2.61 (2.26–2.95)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.90 (1.66–2.15)</td>
<td>2.06 (1.80–2.32)</td>
<td>2.00 (1.68–2.33)</td>
<td>1.81 (1.57–2.05)</td>
<td>2.24 (1.64–2.84)</td>
</tr>
<tr>
<td>France</td>
<td>3.33 (3.10–3.55)</td>
<td>3.51 (3.27–3.74)</td>
<td>3.56 (3.26–3.87)</td>
<td>3.02 (2.75–3.30)</td>
<td>3.87 (3.53–4.20)</td>
</tr>
<tr>
<td>Italy</td>
<td>3.37 (3.18–3.55)</td>
<td>3.55 (3.36–3.75)</td>
<td>3.89 (3.62–4.16)</td>
<td>2.53 (2.23–2.82)</td>
<td>3.48 (3.28–3.67)</td>
</tr>
<tr>
<td>Spain</td>
<td>3.68 (3.49–3.88)</td>
<td>3.89 (3.69–4.09)</td>
<td>4.21 (3.97–4.45)</td>
<td>2.30 (1.94–2.65)</td>
<td>3.94 (3.72–4.15)</td>
</tr>
<tr>
<td>Greece</td>
<td>2.42 (2.23–2.60)</td>
<td>2.59 (2.40–2.78)</td>
<td>3.11 (2.83–3.39)</td>
<td>1.70 (1.49–1.91)</td>
<td>2.60 (2.38–2.83)</td>
</tr>
</tbody>
</table>

Prevalence: % (95% CI)

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Education level</th>
<th>Verbal fluency</th>
<th>Animal naming</th>
<th>Standardised for the 5 effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>19.2 (17.8–20.6)</td>
<td>19.5 (18.1–20.9)</td>
<td>19.2 (17.7–20.5)</td>
<td>19.1 (17.7–20.5)</td>
<td>20.4 (18.8–22.1)</td>
</tr>
<tr>
<td>Denmark</td>
<td>18.1 (16.2–19.9)</td>
<td>18.2 (16.3–20.2)</td>
<td>17.7 (15.8–19.6)</td>
<td>20.3 (18.0–22.7)</td>
<td>18.8 (16.9–20.1)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>19.4 (18.0–20.1)</td>
<td>19.6 (18.1–21.0)</td>
<td>19.5 (18.0–21.0)</td>
<td>18.9 (17.5–20.4)</td>
<td>20.0 (18.5–21.6)</td>
</tr>
<tr>
<td>Germany</td>
<td>18.5 (17.1–19.9)</td>
<td>18.8 (17.4–20.2)</td>
<td>19.0 (17.6–20.5)</td>
<td>23.9 (21.7–26.0)</td>
<td>19.4 (17.9–20.9)</td>
</tr>
<tr>
<td>Austria</td>
<td>19.6 (17.8–21.3)</td>
<td>19.2 (17.5–20.9)</td>
<td>19.7 (17.9–21.5)</td>
<td>22.1 (20.1–24.2)</td>
<td>20.2 (18.4–22.1)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>19.1 (16.6–21.5)</td>
<td>19.2 (16.8–21.7)</td>
<td>18.7 (16.2–21.2)</td>
<td>20.5 (17.4–23.6)</td>
<td>19.7 (17.1–22.4)</td>
</tr>
<tr>
<td>France</td>
<td>33.3 (31.1–35.5)</td>
<td>33.3 (31.0–35.2)</td>
<td>33.3 (31.0–35.6)</td>
<td>33.2 (30.1–35.4)</td>
<td>33.6 (31.3–35.8)</td>
</tr>
<tr>
<td>Italy</td>
<td>33.7 (31.8–35.5)</td>
<td>33.8 (31.8–35.5)</td>
<td>34.3 (32.5–36.2)</td>
<td>29.9 (27.9–32.0)</td>
<td>31.7 (29.8–33.6)</td>
</tr>
<tr>
<td>Spain</td>
<td>36.8 (34.9–38.8)</td>
<td>36.1 (34.3–38.0)</td>
<td>35.9 (34.0–37.9)</td>
<td>30.0 (27.6–32.4)</td>
<td>34.9 (32.9–36.9)</td>
</tr>
<tr>
<td>Greece</td>
<td>24.2 (22.3–26.0)</td>
<td>23.7 (22.2–25.5)</td>
<td>25.1 (23.0–26.9)</td>
<td>22.6 (20.8–24.4)</td>
<td>23.5 (21.7–25.4)</td>
</tr>
</tbody>
</table>

I. Defined as EURO–D score of 4 or above.
for by cognitive impairment (cognitive assessments were not available from most of the EURODEP centres, so this could not be tested directly). This hypothesis is supported in the current analysis, but it is impairment in verbal fluency rather than memory that seems to be mediating or confounding the effect of age on motivation factor scores. As with the EURODEP analyses, the differences between SHARE countries in the distribution of age and gender could not account for between-country differences in depression symptoms. We have further demonstrated that compositional differences in education and cognitive functioning are also not relevant.

Comparisons with other European surveys are limited by the different age ranges and different outcomes studied. For instance, the ESEMeD study (Alonso et al., 2004), as part of the wider World Mental Health survey, used the World Health Organization’s Composite International Diagnostic Interview (CIDI) to estimate the prevalence of mood disorder (DSM-IV bipolar disorder, major depression and dysthymia) in nationally representative samples of all those aged 18 years and over in seven European countries. In descending order, the prevalence of mood disorder varied from Ukraine (9.1%), France (8.5%), The Netherlands (6.9%), Belgium (6.2%), Spain (4.9%), Italy (3.8%) to Germany (3.6%) (Alonso et al., 2004).

The Outcome of Depression International Network (ODIN) study used a two-phase design (the Beck Depression Inventory for screening in the first phase and the Schedule for Clinical Assessment in Neuropsychiatry for definitive clinical diagnoses in the second phase) in locally representative samples from five European countries: the prevalence of any ICD–10 depressive disorder varied from 17.1% in Liverpool (UK) and 12.3% in Dublin (Ireland) to 2.6% in Santander (Spain) (Ayuso-Mateos et al., 2001). Findings from surveys using structured clinical diagnostic assessments of predominately younger adult samples are therefore not consistent with the ethnocultural distribution of reported depression symptoms observed in the SHARE study.

**Cultural and methodological effects on measurement**

We have previously demonstrated, using the SHARE data (Castro-Costa et al., 2007), that the EURO–D had promisingly invariant measurement properties – that is, the factor structures, factor loadings and hierarchical measurement properties were similar across all ten centres. Similar characteristics were observed internationally for the CIDI major depression items in the WHO international study of psychological problems in general healthcare (Simon et al., 2002). The investigators in the latter study remind us that invariant measurement properties do not preclude the possibility of threshold effects, whereby the severity with which a symptom is experienced before the relevant item is endorsed may vary between cultural settings, and that this may account for cultural differences in prevalence (Simon et al., 2002). In other domains of health assessment innovative techniques are being developed to adjust for such effects, using vignettes to estimate and then adjust for threshold differences between populations (Salomon et al., 2004); these could in principle be applied to the assessment of depression. Earlier studies of patterns of responses to items of the Center for Epidemiologic Studies Depression Scale between minority ethnic groups in the USA (Iwata et al., 1995) and in undergraduate students in east Asia and in North and South America (Iwata & Buka, 2002) indicated greater cross-cultural variability in responses to positively worded compared with negatively worded items. Interestingly, the same pattern was observed for the EURO–D in both the EURODEP and SHARE studies. The motivation items (for which there was greater variation in item prevalence and factor scores) are all positively worded, whereas the affective suffering items are negatively worded. Some of this between-country variation may therefore reflect cultural and linguistic differences in appraising and responding to positively worded items, rather than national differences in psychological morbidity. This could not, however, have accounted for the heavier load of depressive symptoms in Latin countries observed in the SHARE study, given that the higher symptom prevalence in France, Italy and Spain was observed for both affective suffering and motivation symptoms (see Figs 1 and 2).

The specific association between verbal fluency (but not memory) and motivation (but not affective suffering) calls into question the construct validity of the motivation components of the EURO–D scale, which may be measuring the effects of subcortical brain damage (apathy and slowing) rather than depression per se.

Given that the EURO–D items were originally selected because they were present in all or most of the five late-life depression assessments used in the EURODEP studies, this is likely to be a general problem. It would be interesting to re-explore the concept of ‘vascular depression’ (Alexopoulos et al., 1997), using the affecting suffering and motivation components. It is tempting to conclude that the affective suffering component might be the more valid measure of psychological morbidity, as well as providing a more psychometrically appropriate tool for cross-cultural comparison (Castro-Costa et al., 2007).

**Concluding remarks**

Given the pattern of findings, it is tempting to conclude that variation in prevalence of depressive symptoms and syndromes in older European men and women may be best understood in terms of ethnocultural differences, with a higher prevalence recorded in the Latin nations (France, Italy and Spain) than in the Germanic (Sweden, Denmark, Germany, The Netherlands) and Hellénic (Greece) countries. However, although we have excluded here the compositional effects of major determinants of depression prevalence – age, gender, education and cognitive function – it remains possible that other risk exposures, differently distributed between countries, might have accounted for the observed variation. Even though compositional effects can be confidently excluded, it remains difficult to attribute the contextual effect or effects that might be responsible. Language and culture are contextual effects, in that they are the property of the population (country, in this case) with no meaningful individual-level variation. Other contextual effects may be important, for example income inequality (Muramatsu, 2003), social capital (LaGory, 1992) and religiosity (Braam et al., 2001). Technically it is possible in principle to study contextual effects and to disentangle their impact from that of compositional effects, using multilevel modelling approaches. This will be the focus of a further analysis.

**ACKNOWLEDGEMENTS**

The SHARE survey was funded primarily by the European Commission through the Fifth Framework Programme (project QLK6–CT-2001-00360 in the thematic programme Quality of Life). Additional funding came from the US National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OIGHA.
04–064). Data collection in Austria (through the Austrian Science Fund, FWF), Belgium (through the Belgian Science Policy Office) and Switzerland (through BBW/OFES/UFES) was funded within those countries. E.C.-C. was supported for this analysis by the Social Psychiatry Research Trust.

REFERENCES


Genotype effects of CHRNA7, CNRI and COMT in schizophrenia: interactions with tobacco and cannabis use

STANLEY ZAMMIT, GILLIAN SPURLOCK, HYWEL WILLIAMS, NADINE NORTON, NIGEL WILLIAMS, MICHAEL C. O'DONOVAN and MICHAEL J. OWEN

Background Genetic variations might modify associations between schizophrenia and cannabis or tobacco use.

Aims To examine whether variants within the cannabinoid receptor (CNRI) and \(\alpha_2\) nicotinic receptor (CHRNA7) genes are associated with schizophrenia, and whether these effects vary according to cannabis or tobacco use. We also examined a putative interaction between cannabis and Val\textsuperscript{158}Met within the catechol-O-methyltransferase gene (COMT).

Method Genotype effects of CHRNA7 and CNRI were studied in a case–control sample of 750 individuals with schizophrenia and 688 controls, with interactions for these genes studied in small subsamples. A case–only design of 493 of the schizophrenia group was used to examine interactions between cannabis use and COMT.

Results There was no evidence of association between schizophrenia and CNRI (OR = 0.97, 95% CI 0.82–1.13) or CHRNA7 (OR = 1.07, 95% CI 0.77–1.49) genotypes, or of interactions between tobacco use and CHRNA7, or cannabis use and CNRI or COMT genotypes.

Conclusions Neither CNRI nor CHRNA7 variation appears to alter the risk of schizophrenia. Furthermore, our results do not support the presence of different effects of cannabis use on schizophrenia according to variation within COMT.

Declaration of interest None.

Schizophrenia is associated with increased use of cannabis and tobacco compared with the general population, although reasons for these associations have not been clearly elucidated. There is some evidence that people with schizophrenia may use tobacco to alleviate neurophysiological deficits associated with this disorder (Adler et al., 1993; Olincy et al., 1998), and that this is mediated through effects at the \(\alpha_2\) nicotinic acetylcholine receptor (CHRNA7) (Gray et al., 1996; Stevens et al., 1998). An association between schizophrenia and a putative functional variant, ~86C/T, within the CHRNA7 gene (CHRNA7) has been reported (Leonard et al., 2002) and warrants further exploration.

The main psychoactive compound within cannabis is delta-9-tetrahydrocannabinol (\(\Delta^2\)-THC), which acts through the CNRI cannabinoid receptor. An increased incidence of psychotic disorders in people using cannabis has been observed (Arseneault et al., 2002; Zammit et al., 2002) and a putative interaction between cannabis use and variation within the catechol-O-methyltransferase (COMT) gene on risk of psychosis has also been reported (Caspi et al., 2005). Findings from relatively small studies examining association between CNRI genetic variation – most commonly at the single nucleotide polymorphism (SNP) rs1049335 – and schizophrenia have been inconsistent, and it was considered worth while to examine this in a substantially larger sample than has been studied thus far.

The main aims of our study were to investigate whether variations at ~86C/T within CHRNA7 and at rs1049335 within CNRI were associated with schizophrenia, and whether these relationships differed according to use of tobacco or cannabis. We also investigated whether there was any evidence of an interaction between cannabis use and the Val\textsuperscript{158}Met polymorphism (SNP rs4680) within COMT, as previously reported (Caspi et al., 2005), as well as with SNPs rs737865 and rs165599 within this gene. The SNP rs4680 alters enzyme activity of COMT (Chen et al., 2004), whereas the GGG haplotype of SNPs rs737865–4680–165599 has been reported to be associated with lower expression of COMT messenger RNA in human brain tissue (Bray et al., 2003) and with increased risk of schizophrenia (Shifman et al., 2002). Main genotype and haplotype effects of COMT in this sample have been previously reported, with no evidence found for any association with schizophrenia (Williams et al., 2005).

METHOD

Participants A sample of unrelated individuals with schizophrenia was recruited from outpatient and in-patient clinical settings and from volunteer support organisations within the UK. These individuals were assessed using the Schedule for Assessment of Neuropsychiatric Disorders semi-structured interview (SCAN; Wing et al., 1990) together with case-note review wherever possible. The Operational Criteria Checklist (OPCRIT; McGuffin et al., 1991) and Global Assessment Scale (GAS; Endicott et al., 1976) were also completed. High levels of reliability (\(k > 0.8\)) were achieved between raters for diagnoses and rating scale items. Controls were unrelated blood donors ascertained from the same regions as the majority of the patients. Given the prevalence of schizophrenia and the fact that people taking regular medication cannot be blood donors in the UK, it was not deemed necessary to screen the control group for schizophrenia to retain statistical power (Owen et al., 1997). Ethical approval was granted for this study and informed consent was obtained from all participants.

All study participants were White, with both parents born in the UK or Ireland. All cases of schizophrenia satisfied DSM-IV criteria (American Psychiatric Association, 1994) for consensus lifetime diagnosis of the disorder, made by two independent raters. The following phenotypes, determined a priori, were examined in relation to ~86C/T and rs1049335 genotype:

(a) age at onset, defined as the age at which psychiatric help for psychotic symptoms was first sought;

(b) worst-ever GAS score, ranging from 0 (most severe) to 100 (least severe).
Genotyping

The CHRNA7 promoter polymorphism –86C/T was genotyped as a restriction-fragment length polymorphism using the restriction enzyme Hph1 (New England Biolabs, Ipswich, Massachusetts, USA). The primers were 5′-agtagctccgacactccctc-3′ and 5′-atgttcgagccgacgctg-3′ as used by Leonard et al. (2002). The product was amplified using the GC-RICH PCR System (Roche Diagnostics, Basel, Switzerland), and the 272 base pairs fragment was digested with Hph1 resulting in two fragments of 79 bp and 193 bp with the T allele. The products were run out on a 1.5% agarose gel and visualised using ethidium bromide.

The CNR1 polymorphism rs1049353 was genotyped by fluorescence polarisation using an AcycloPrime kit (PerkinElmer, Waltham, Massachusetts, USA) and the output was read on an IJL Biosystems (Sunnyvale, California, USA) plate reader. A 297 bp amplimere was amplified using primers 5′-tccctctggaagacat-3′ and 5′-tatgagcatggttaaatc-3′. The SNP was at position 125. The extension primer used was an AcycloPrime kit (PerkinElmer, Boston, Massachusetts, USA).

In the fluorescence polarisation assay was 5′-catactcctggaacatgc-3′. The COMT markers were genotyped using SNaPshot (Applied Biosystems, Foster City, California, USA) using an ABI3100 sequencer. Details of primers and reaction conditions are provided in Appendix 1 at http://www.cardiff.ac.uk/medicine/psychological_medicine/pub_data/comt.htm.

Analysis

The reference participants for the analyses were those with genotypes that were CC homozygous for –86C/T, GG homozygous for rs1049353, AA homozygous for rs737865, AA homozygous for rs165599, and homozygous for the Met allele at Val158Met within COMT. Only 0.4% of our participants were homozygous for the Met allele at the –86C/T locus, and they were therefore grouped with the C/T heterozygotes.

Logistic regression was used to examine associations between dichotomous outcomes and genotypes. A dominance genetic model, as described above, was examined for –86C/T, whereas additive models were used for the CNR1 and COMT variants (Lewis, 2002). For the study of continuous phenotypic outcomes, linear regression was used. However, for age at onset, where assumptions of normality were not met, data were In-transformed prior to regression modelling. Statistical interactions on a multiplicative scale between substance use and genotype on risk of schizophrenia were investigated using a likelihood ratio test within the logistic regression models. For Val158Met, however, as no association was observed between this SNP and cannabis use in the Dunedin cohort (Caspi et al., 2005), we used a case-only approach to investigate possible gene-environment interactions because this is statistically more powerful (Khoury & Flanders, 1996). The case-only analysis was also used for rs737865 and rs165599 within COMT. Haplotypes for COMT were examined using UNPHASED, version 3.0 (Dudbridge, 2003).

This study had greater than 95% power to detect an additive genetic effect with an odds ratio of 1.4 or above at x=0.05 for the CNR1 and COMT variants examined. This study also had greater than 95% power to find an association between –86C/T variation and schizophrenia based on frequencies of CC genotype of 0.91 in the control group and 0.84 in the schizophrenia group, as observed by Leonard et al. (2002). The interaction odds ratio previously reported for cannabis and Val158Met was 3.5 (Caspi et al., 2005), and our case-only approach had more than 90% power to detect an interaction odds ratio of as low as 1.5, at x=0.05.

Sensitivity analysis

Some participants were likely to have started using tobacco or cannabis after the onset of schizophrenia and it is possible that this could obscure and complicate interpretation of results from this study. Examination of the association between schizophrenia and genotypes was therefore repeated with analyses restricted to cases where the onset of substance use was reported to be at least 1 year prior to age at schizophrenia onset.

RESULTS

There were 838 participants with schizophrenia who were genotyped for any of CNR1 (n=797), CHRNA7 (n=750) or COMT (n=575). Data on cannabis and tobacco use were missing for 96 (11.5%) and 107 (12.8%) of these respectively. Of those with substance use data, 276 participants (37.2%) had ever used cannabis, and 531 (72.6%) had ever used tobacco.

CHRNA7

The –86C/T genotypes were in Hardy–Weinberg equilibrium in both the schizophrenia group (χ²=0.01, P=0.76) and the control group (χ²=0.91, P=0.92). As shown in Table 1, there was no evidence for any association between –86C/T genotype and schizophrenia (CT/TT genotypes OR=1.07, 95% CI 0.77–1.49; P=0.70). There was little evidence of any difference in the effect of genotype on schizophrenia between those who smoked (schizophrenia group n=473, controls n=24; OR=3.0, 95% CI 0.4–22.9) and those who did not (schizophrenia group n=186, controls n=25; OR=1.7, 95% CI 0.4–7.7; interaction likelihood ratio test χ²=0.21, d.f.=1, P=0.65). As tobacco use data were available only for a small proportion of the control group, a more powerful case-only analysis was also used, and this also failed to provide any evidence for interaction (n=659; odds ratio for tobacco use by CHRNA7 genotype 0.89, 95% CI 0.53–1.48).

There were 123 in the schizophrenia group with data relating to age of first using tobacco, and 104 (85%) of these
claimed to have started using tobacco prior to the onset of schizophrenia. In the sensitivity analysis there was similarly little evidence of any difference in the effect of genotype on schizophrenia between non-smokers and those smoking prior to illness onset (n=110; OR=2.7, 95% CI 0.3–22.3; interaction \( \chi^2=0.1, \text{d.f.}=1, P=0.73 \)).

Another way of presenting these data is to examine the relationship between tobacco use and schizophrenia stratified by –86C/T genotype. Tobacco use was strongly associated with schizophrenia in the whole sample (OR=4.4, 95% CI 3.3–6.0; P<0.001), with no evidence of any interaction when stratified by genotype (CC genotype OR=2.6, 95% CI 1.4–4.7; CT/TT genotypes OR=4.6, 95% CI 0.4–33.0; interaction likelihood ratio test as above, P=0.65). Tobacco use was not associated with –86C/T genotype (OR=0.9, 95% CI 0.5–1.5).

Results for associations between –86C/T genotype and various phenotypes within schizophrenia are presented in Table 2. There was weak evidence (P=0.07) that participants with the CT/TT genotypes had a younger age of onset, by approximately 2 years on average, than those homozygous for the C allele.

**CNRI**

Genotypes at rs1049353 were in Hardy–Weinberg equilibrium in both the schizophrenia group (\( \chi^2=0.56, \text{P}=0.44 \)) and controls (\( \chi^2=1.0, \text{P}=0.36 \)). As shown in Table 1, there was no evidence for any association between rs1049353 genotype and schizophrenia (odds ratio for linear trend of genotypes 0.97, 95% CI 0.82–1.13; P=0.66).

There was little evidence of any difference in the effect of rs1049353 genotype on schizophrenia between those who did not use cannabis (schizophrenia group n=445, controls n=93; OR=1.04, 95% CI 0.73–1.47) and those who did (schizophrenia group n=261, controls n=23; OR=0.92, 95% CI 0.48–1.75; interaction \( \chi^2=0.11, \text{d.f.}=1, P=0.74 \)). As cannabis use data were again available for only a small proportion of the control group, a case-only analysis was used, and this also failed to provide any evidence for interaction (n=706; odds ratio for cannabis use by CNRI genotype 0.83, 95% CI 0.65–1.05).

As part of the sensitivity analysis, there were 71 individuals in the schizophrenia group with data relating to age of first use in our sample of 493 persons with schizophrenia (OR=0.98, 95% CI 0.76–1.27, P=0.89). Results were almost identical when restricting the analysis to participants who first used cannabis at least 1 year prior to their illness onset and who had first used it by age 18 years or earlier (n=338; OR=0.76, 95% CI 0.41–1.40; P=0.38). Similarly, there was no evidence that variation at rs737865 or rs165599 was associated with cannabis use in the case-only analysis, even when restricting the analysis to first use of cannabis at least 1 year prior to illness onset and first use by age 18 years or earlier (rs737865, OR=1.09, 95% CI 0.56–2.00; rs165599, OR=1.09, 95% CI 0.57–2.08). There was no evidence of overall haplotype association with cannabis use (\( \chi^2=4.7, \text{d.f.}=7, \text{P}=0.69 \)) or of specific association with the rs737865–4680–165599 GGG haplotype (\( \chi^2=0.001, \text{d.f.}=1, \text{P}=0.98 \)).

**DISCUSSION**

There was no evidence for any association between CHRNA7 or CNRI genotype and schizophrenia in our sample, and also

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Association between CHRNA7 (–86C/T) and CNRI (rs1049353) genotypes and schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number (%) by genotype</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------</td>
</tr>
<tr>
<td>–86C/T</td>
<td></td>
</tr>
<tr>
<td>Control group</td>
<td>548 (88.7)</td>
</tr>
<tr>
<td>Schizophrenia group</td>
<td>660 (88.0)</td>
</tr>
<tr>
<td>rs1049353</td>
<td>GG</td>
</tr>
<tr>
<td>Control group</td>
<td>335 (48.7)</td>
</tr>
<tr>
<td>Schizophrenia group</td>
<td>407 (51.1)</td>
</tr>
</tbody>
</table>

1. In CT/TT vs CC.
2. Per A allele.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Effect estimates for phenotype characteristics according to CHRNA7 (–86C/T) and CNRI (rs1049353) genotypes in participants with schizophrenia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β (95% CI)</td>
</tr>
<tr>
<td>Age at onset¹</td>
<td>–0.06 (–0.13 to 0.01)</td>
</tr>
<tr>
<td>GAS score</td>
<td>–0.9 (–2.7 to 0.9)</td>
</tr>
<tr>
<td>Paranoid delusions</td>
<td>–0.05 (–0.3 to 0.2)</td>
</tr>
<tr>
<td>Disorganised symptoms</td>
<td>0.11 (–0.1 to 0.4)</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>0.01 (–0.2 to 0.3)</td>
</tr>
<tr>
<td>First-rank delusions</td>
<td>–0.01 (–0.3 to 0.2)</td>
</tr>
<tr>
<td>Poor outcome</td>
<td>OR=1.30 (0.8 to 2.1)</td>
</tr>
</tbody>
</table>

GAS, Global Assessment Scale.
1. Effect estimate is comparing CC with CT/TT genotypes for –86C/T.
2. Effect estimate is per A allele for CNRI.
3. In transformed.
no evidence of any gene–environment interactions between tobacco use and CHRNA7, or cannabis use and CNR1 or COMT genotype.

**CHRNA7 and tobacco use**

There have been few association studies of polymorphisms within CHRNA7 and schizophrenia to date. Leonard et al (2002) screened the core promoter region of the full-length gene and reported an association between schizophrenia and variant –86C/T. Although we found no evidence for an association between the promoter SNP –86C/T and schizophrenia, CT/TT genotypes occurred slightly more frequently in participants with schizophrenia than in controls, in a direction consistent with the findings by Leonard et al (2002). However, we observed a much smaller difference in CC frequency of less than 1%, as opposed to the 7% reported in the original study (Leonard et al, 2002).

People with schizophrenia commonly display evidence of sensory attention impairments (Adler et al, 1982; Leonard et al, 1996), including deficits in pre-pulse inhibition and P50 gating response (Braff & Saccuzzo, 1985; Braff et al, 1992; Waldo et al, 1994). Improvements in such neurophysiological deficits in people with schizophrenia following cigarette smoking have been reported (Adler et al, 1993; Olincy et al, 1998), with similar improvements observed following nicotine administration in animal models (Bickford & Wear, 1995; Stevens et al, 1996, 1998). Specific agonists of the α7 receptor (CHRNA7) normalise sensory gating deficits in animal models (Stevens et al, 1998), whereas evidence for genetic linkage to the P50 deficit and, to a lesser extent, to schizophrenia, has been reported for chromosome band 15q14, an area that contains CHRNA7 (Coon et al, 1993; Freedman et al, 1997; Leonard et al, 1998).

Despite this support, from a variety of sources, that CHRNA7 is a good candidate gene for schizophrenia, there is weak evidence at present that variation within this gene is associated with the disorder (Riley et al, 2000; Xu et al, 2001; Leonard et al, 2002; Gault et al, 2003; Li et al, 2004; Fan et al, 2006). However, given the findings from experimental studies of the effect of nicotine on neurophysiological deficits in both animal models and humans, as described earlier, it may be that any association between CHRNA7 and schizophrenia is mediated by impairments in sensory gating or other related physiological responses. In the study by Leonard et al (2002), presence of the T allele at –86C/T was also associated with reduced inhibition of the P50 response in the control group, and although two other studies did not replicate this finding (Gault et al, 2003; Houy et al, 2004), one reported an association between P50 sensory gating response and another promoter variant, –194G/C (Houy et al, 2004). There is a clear need for research into CHRNA7 variation in relation to neurophysiological deficits in well-designed and adequately powered studies to address this further.

**CNR1, COMT and cannabis**

We found no evidence of association between the CNR1 locus rs1049353 and schizophrenia, consistent with the overall findings previously reported for this variant from two much smaller studies (Leroy et al, 2001; Ujike et al, 2002), although one of these reported an association in a subgroup analysis (Leroy et al, 2001). Two studies have reported associations between schizophrenia and variation within an (AAT)\textsubscript{n} microsatellite approximately 20kb upstream of the translational start site of CNR1 (Ujike et al, 2002; Martinez-Gras et al, 2006). However, different alleles were associated with increased risk in these two studies, and the association in one of the studies was again observed only for a subgroup of participants, this time with hebephrenic schizophrenia.

The CNR1 gene is located on 6q14–15, a region of replicated linkage for schizophrenia (Lewis et al, 2003). There are four SNPs within CNR1 on HapMap that have a heterozygosity in European populations greater than 0.1; three of these are in the 3’ untranslated region whereas rs1049353 is a synonymous SNP found within exon 1. The relatively small size of CNR1, the limited variation within the gene and its linkage disequilibrium structure mean it is unlikely that a substantial effect on schizophrenia risk is conferred by variation within this gene, given our findings and the lack of other consistent associations reported to date.

We also failed to find any supporting evidence for a differential effect of cannabis use on psychosis risk according to variation at Val\textsuperscript{10}Met within COMT. In the Dunedin study evidence for an interaction was observed only for people first using cannabis by age 18 years, but not for those using it after this age (Casp et al, 2005). One explanation proposed for this was that there may exist a sensitive or even critical period of risk when the influence of cannabis exposure is moderated by COMT genotype. In our study we failed to find evidence for an interaction between cannabis use and COMT genotype even when restricting the analysis to participants who claimed to have first used cannabis by the same cut-off period of age 18 years, despite more than adequate statistical power to replicate the original findings. Furthermore, in contrast to the findings by Caspi et al (2005), cannabis use by age 18 years was actually less common in participants with schizophrenia homozygous for the Val allele compared with those heterozygous for this allele or homozygous for Met (5.3%, 6.4% and 8.7% respectively), although this was not significantly different.

**Limitations to the interpretation of our results**

This study was adequately powered to examine main effects on risk of schizophrenia, suggesting it is unlikely that variations in CNR1 or CHRNA7 are important risk factors for schizophrenia. Furthermore, this study was adequately powered for studies of interactions using a case-only design, but this approach is dependent on the assumption of no genotype–exposure association in the population. For COMT this assumption is likely to be a reasonable one, given that no association with cannabis use was observed in the Dunedin cohort (Casp et al, 2005). However, this assumption may be less likely to hold true for CNR1 or CHRNA7, given that cannabis and nicotine act through receptors coded for by these genes, and also given the sporadic reports of associations between cannabis and tobacco dependence and CNR1/CHRNA7 genotypes (Greenbaum et al, 2006; Hopfer et al, 2006). For that reason we also conducted studies of interactions between CNR1 and cannabis as well as between CHRNA7 and tobacco using a more traditional case–control approach, although statistical power to exclude anything other than large interaction effects for these two genes using this latter approach was limited.

Although we genotyped three SNPs in COMT that together form a haplotype reported to be significantly associated with schizophrenia (Shiftman et al, 2002), we
only genotyped one SNP in each of CNR1 and CHRNA7. It is not possible therefore to rule out causal effects of variants within these genes that are not in strong linkage disequilibrium with the SNPs we tested. However, a strong effect of CNR1 on risk of schizophrenia seems unlikely, given the linkage disequilibrium structure within this gene. Our confidence in ruling out such an effect for CHRNA7 is lower, although we did not feel that the evidence we obtained was strong enough to warrant further genotyping of CHRNA7 SNPs, especially given the problems resulting from the partial duplication of this gene, which makes such studies inherently more difficult.

A final limitation of our study is that, unlike the longitudinal data collection in the Dunedin cohort, our case–control design relied on people recalling age of first use of cannabis and relating this in time to the date of their first contact with psychiatric services. Such data seem inherently more likely to be misclassified than prospectively collected data. It is unclear to what extent any such misclassification might have resulted in an underestimate of the association between cannabis use and genotype in our case-only analysis, and therefore obscured any true interaction effect. It would, however, presumably require a substantial amount of misclassification to obscure an interaction effect as strong as that reported by Caspi and colleagues, whereby cannabis use was associated with a 10-fold increase in risk of psychotic disorder in those homozygous for valine but had no effect in those homozygous for methionine (Caspi et al., 2005). This finding of an interaction effect in the Dunedin cohort was observed only in a subgroup of participants—those using cannabis by age 18 years. Similarly, supportive evidence of a putative interaction between cannabis use and COMT on psychotic symptoms, following administration of cannabis in an experimental setting was again observed only in a subgroup of participants with schizophrenia, this time those with evidence of pre-existing psychotic traits (Henquet et al., 2006). Although such findings are biologically plausible and seem intuitively appealing, substantially more evidence from replication of these findings is required. Our study, although providing adequate power to observe even a relatively small association between cannabis use and COMT genotype in participants with schizophrenia, may not be the ideal design to examine such a relationship, and other longitudinal studies may be able to investigate this with greater confidence in the future.

In summary, we failed to find any evidence that variation at the CHRNA7 or CNR1 locus was associated with schizophrenia, or that the effect of variation at these loci was modified by use of tobacco or cannabis respectively. Cannabis use was not associated with presence of the valine allele at Val158Met within COMT in our sample, therefore our findings do not support a previous report of a putative gene–environment interaction between COMT genotype and cannabis use on risk of schizophrenia.

ACKNOWLEDGEMENTS

S.Z. is funded through a Clinician Scientist Award from the National Assembly for Wales.

REFERENCES


Psychophysiological and behavioural characteristics of individuals comorbid for antisocial personality disorder and schizophrenia-spectrum personality disorder

ROBERT A. SCHUG, ADRIAN RAINE and RAND R. WILCOX

Background Few studies have examined people with comorbid schizophrenia-spectrum personality disorder and antisocial personality disorder, a subgroup who may differ psychophysically and behaviourally from those with either condition alone.

Aims To test whether individuals with both types of personality disorder are particularly characterised by reduced orienting and arousal and by increased criminal offending.

Method In a community adult sample, self-reported crime and skin conductance orienting were collected on four diagnostic groups: schizophrenia-spectrum personality disorder only; antisocial personality disorder only; comorbidity of the two disorders; and a control group.

Results The comorbid group showed significantly higher levels of criminal behaviour than the other three groups. They also showed reduced skin conductance orienting to neutral tones compared with the other groups, and significantly reduced arousal and orienting to significant stimuli compared with the control group.

Conclusions Reduced orienting may reflect a neurocognitive attentional risk factor for both antisocial and schizotypal personality disorders that indirectly reflects a common neural substrate to these disorders.

Declaration of interest None. Funding detailed in Acknowledgements.

A second generation of schizophrenia and crime research – focusing upon aetiological processes predisposing to both conditions – may be advanced by examining the comorbidity between schizophrenia-spectrum personality disorders (paranoid, schizoid and schizotypal) and antisocial personality disorder, in an attempt to ascertain if people with this comorbidity differ in biological and/or behavioural ways beyond the additive effects of either condition alone. Although studies suggest that people with such comorbidity demonstrate skin conductance orienting response abnormalities (Raine, 2006a) and are more antisocial than those with antisocial personality disorder alone (Raine et al, 1999; Moran & Hodgens, 2004), electrodermal arousal deficits in antisocial populations (Fowles, 1993; Lorber, 2004), and mixed arousal findings in samples of people with schizophrenia or schizotypal disorder (Ohman, 1981; Dawson et al, 1994; Raine et al, 2002b) indicate that ignoring comorbidity of schizophrenia-spectrum and antisocial personality disorders may obfuscate findings when examining each condition separately. This study sought to identify autonomic and behavioural characteristics of this comorbid group in the community, and tested the hypotheses that individuals with comorbidity of these disorders would demonstrate increased criminality and/or reduced skin conductance orienting and arousal when compared with individuals with either disorder alone and with controls.

Diagnostic measures Participants were administered the Structured Clinical Interviews for DSM-IV Axis I Disorders (SCID-I; First et al, 1997b) and Axis II Personality Disorders (SCID–II; First et al, 1997a). The SCID-I and SCID-II were administered by a clinical PhD graduate student who had received systematised training in diagnostic assessment (Ventura et al, 1998). The sample was classified into four groups, based upon SCID–II diagnoses and availability of self-report crime and skin conductance data (see below). The group with schizophrenia-spectrum personality disorder (SSPD) contained participants diagnosed with paranoid, schizoid and/or schizotypal paranoid personality disorders, with no comorbid antisocial personality disorder (n=9). The antisocial personality disorder (ASPD) group contained participants with an ASPD diagnosis but no comorbid SSPD (n=14). The comorbid SSPD/ASPD group contained participants who met diagnostic criteria for both SSPD and ASPD (n=8). The control group contained participants with no Axis II diagnosis (n=48). Participants with neither ASPD nor SSPD but with other Axis II psychopathology (n=22) were excluded from group assignment. Groups did not significantly differ on age, gender, ethnicity, IQ or socio-economic status (Hollingshead, 1979) (see Table 1).

Criminal offending Two forms of criminal offending data were collected. First, criminal records were searched and assessed for total numbers of arrests and convictions for each participant (Raine et al, 2000). Second, each participant was administered an adult extension (Raine et al, 2000) of the National Youth Survey self-report delinquency measure (Elliot et al, 1983). Self-report criminal offending was operationalised as the total...
number of property, violence and drug offences assessed by this 50-item instrument.

To encourage open reporting, a certificate of confidentiality was obtained from the Secretary of Health, pursuant to section 303(a) of Public Health Act 42. Participants were informed that any information they might provide about uninvestigated crimes could not be subpoenaed by any US federal, state or local court. Participants were reminded of confidentiality during administration of the measure and the limits to the confidentiality certificate.

**Psychophysiological measures**

Skin conductance was measured during both rest and orienting conditions.

**Apparatus and recording procedures**

Participants were tested in a temperature-controlled, light- and sound-attenuated psychophysiological recording laboratory. Skin conductance was recorded from the distal phalanges of the first and second fingers of both hands (to maximise responsiveness; Scerbo et al., 1992) using Beckman silver-silver chloride electrodes (1 cm diameter) with sodium chloride 0.9% solution in Unibase as electrolyte, with the skin contact area delineated using double-sided adhesive masks with a hole of 1 cm diameter. Recordings were made using a Grass Model 7 polygraph (Grass Instruments, Quincy, Massachusetts, USA) with a constant 0.5 V potential across electrodes to allow direct recording of skin conductance (Venables & Christie, 1980). Participants were made as comfortable as possible and asked to keep their hands completely still. They were then instructed that after a 3 min rest period they would hear a series of tones that would last about 5 min. The amount of time that elapsed from the end of electrode placement to the start of the rest period was approximately 2 min.

**Stimuli**

A set of ten orienting stimuli were presented with inter-stimulus intervals randomised between 25 s and 40 s. Orienting stimuli consisted of a series of six 75 dB tones of 1000 Hz frequency, 25 ms rise time and 1 s duration. These were followed by four more attentionally meaningful stimuli consisting in order of presentation of a re-orienting stimulus (a 500 Hz tone of 75 dB intensity, 1 s duration), a consonant–vowel stimulus ('da', 0.35 s duration, 75 dB intensity), one 90 dB stimulus (1 s duration, 1000 Hz frequency) and one 90 dB white noise burst (1 s duration, 5 ms rise time).

**Scoring**

Skin conductance responses to each orienting stimulus were defined as increases in conductivity of more than 0.05 mS occurring within a latency window of 1–3 s post-stimulus. The number of non-specific skin conductance responses occurring during the 3 min rest period (using the same amplitude criterion as above) was also scored. Skin conductance levels (in mS) were recorded at the beginning and end of the rest period, and at end of the orienting procedure. Charts were scored with the rater masked to group membership. Amplitudes were subjected to square root transformation to reduce skew and kurtosis, as recommended by Venables & Christie (1980). Because levels for the right and left hands were strongly correlated with each other at the respective time points (correlations ranged from 0.77 to 0.83, P < 0.001), the values for the two hands at each time point were averaged. If data were missing for one hand (which occurred on five occasions owing to equipment failure), then data for the available hand were used.

**Statistical analyses**

In cases in which serious violations of the assumptions underlying traditional statistical techniques were detected, additional modern methods were used. It is known that conventional methods for comparing means can have very poor power (e.g. Wilcox, 2005). Comparing medians can reduce problems associated with methods for comparing means, but a concern about using medians is that they trim too much of the data, which again can result in relatively poor power. By trimming 20%, poor power due to outliers, skewness and variance can be reduced substantially, yet good power is still achieved under standard assumptions (Wilcox, 2005). The employment of bootstrapping techniques (see Wilcox, 2003) also seemed appropriate because of the small sizes of some of the groups in our study. Rather than assume normality to determine appropriate critical values, bootstrap methods estimate appropriate critical values using the available data (Efron & Tibshirani, 1993; Davison & Hinkley, 1997).

**RESULTS**

The sample’s demographic characteristics are summarised in Table 1.

**Comorbidity**

Chi-squared analysis was used to assess comorbidity between antisocial personality disorder and schizophrenia-spectrum personality disorder within the entire community sample. Results confirmed significant comorbidity between the two conditions (\(\chi^2 = 7.665, P = 0.006\)). Among those diagnosed with schizophrenia-spectrum spectrum disorder (n=17), almost half (8, 47%) had a comorbid diagnosis of antisocial personality disorder compared with
17% \((n=8)\) in those without schizophrenia-spectrum disorder \((n=84)\) (Fig. 1). Of the total sample, 48 (47.5%) were diagnosed with no Axis II condition, 14 (13.9%) with antisocial personality disorder only, 9 (8.9%) with schizophrenia-spectrum disorder only and 8 (7.9%) with comorbid schizophrenia-spectrum/antisocial personality disorder.

**Criminal offending**

One-way analysis of variance (ANOVA) was used to assess differences in self-reported criminal offending among the four groups. Results indicated that the groups differed significantly: \(F(3)=15.687, P<0.001\). *Post hoc* (Bonferroni-corrected) tests indicated, as expected, that the ASPD group \((n=13);\) mean=21.62, \((s.d.=12.45)\) reported significantly more criminal offending than controls \((n=46);\) mean=11.34, \((s.d.=5.61), P=0.004\). Furthermore, the SSPD/ASPD group \((n=8);\) mean=33.75, \((s.d.=17.37)\) reported significantly more criminal offending than the ASPD group \((P=0.026)\), the SSPD group \((n=8);\) mean=14.25, \((s.d.=9.00), P<0.001\) and the control group \((P<0.001)\). A boxplot of the self-report crime data indicated only minimal outliers and skewness. Consequently, additional modern bootstrapping methods involving trimmed means were not applied to augment this specific analysis.

Additional ANOVAs were used to assess differences in both the total number of charges and convictions among the four groups. Results indicated that the criminal records of the SSPD/ASPD group contained more charges and convictions than those of the other groups (Table 2); however, the groups did not differ significantly on the total number of either charges \((F(3)=1.183, P=0.322)\), or convictions \((F(3)=1.181, P=0.323)\). Boxplots of these data indicated significant outliers in some cases, thus the aforementioned modern methods were employed to augment conventional analyses. A percentile bootstrap method for 20% trimmed means indicated that the criminal records of both the ASPD group and the SSPD/ASPD group contained significantly more charges than the SSPD group and the controls \((P<0.0005\) and \(P<0.0001\) respectively for the ASPD group; \(P=0.006\) and \(P=0.003\) respectively for the SSPD/ASPD group). Although the SSPD/ASPD group had 97% more charges than the ASPD group, this difference was not statistically significant.

An additional percentile bootstrap method for 20% trimmed means indicated that the criminal records of both the ASPD group and the SSPD/ASPD group contained significantly more convictions than the SSPD group and the controls \((P=0.0005\) and \(P=0.001\) respectively for the ASPD group; \(P=0.006\) and \(P=0.007\) respectively for the SSPD/ASPD group). Although the SSPD/ASPD group showed a 98% increase in convictions compared with the ASPD group, given the modest sample size this large difference was not statistically significant.

**Psychophysiology**

**Skin conductance arousal**

A conventional repeated-measures multivariate analysis of variance (MANOVA) was used to assess group differences in skin conductance level during testing (i.e. at the beginning and end of the initial 3 min rest period and at the end of the orienting phase). Results indicated that the groups differed significantly on the three skin conductance level readings; \(F(3)=3.224, P=0.027\). *Post hoc* tests, however, indicated that comparisons were largely non-significant after Bonferroni correction, and that only the SSPD/ASPD group demonstrated a trend toward significantly lower skin conductance arousal when compared with the control group \((P<0.069)\). Boxplots of these data indicated relatively few outliers but significant skewness in some cases. A repeated-measures bootstrap analysis confirmed that the SSPD/ASPD group demonstrated significantly lower skin conductance level than controls \((P<0.001)\) and that all other group differences remained non-significant.

**Skin conductance responsivity**

**Frequency.** An ANOVA was used to assess the difference in total number of skin conductance responses during the entire orienting phase of the procedure. Additional ANOVAs were used to assess differences in the total number of responses during the first six orienting stimuli only, and then during the four meaningful stimuli only. Boxplots of these data indicated relatively few outliers but significant skewness in some cases. Consequently, modern methods were employed to augment conventional analyses.

Results indicated that the groups differed significantly in number of skin conductance orienting responses during the entire orienting phase: \(F(3)=4.104, P=0.009\). *Post hoc* tests (Bonferroni-corrected) indicated that only the SSPD/ASPD group \((n=8);\) mean=0.500, \((s.d.=0.598)\) demonstrated significantly fewer responses over the entire orienting phase than the control group \((n=48);\) mean=2.583, \((s.d.=2.025), P=0.024\). A bootstrap test of linear contrasts using 20% trimmed means, however, indicated that both the ASPD and the SSPD/ASPD groups demonstrated significantly fewer responses over the entire orienting phase than controls \((P=0.027\) and \(P=0.002\) respectively). Conventional ANOVA results also indicated that the groups did not differ significantly in number of responses during the first six orienting

---

**Table 2** Criminal offending for the four diagnostic groups

<table>
<thead>
<tr>
<th></th>
<th>SSPD Mean (s.d.)</th>
<th>ASPD Mean (s.d.)</th>
<th>SSPD/ASPD Mean (s.d.)</th>
<th>Controls Mean (s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported crime</td>
<td>14.25 (9.00)</td>
<td>21.62 (12.45)*</td>
<td>33.75 (17.38)**</td>
<td>11.35 (5.61)</td>
</tr>
<tr>
<td>Charges</td>
<td>0.44 (0.88)</td>
<td>4.57 (4.00)*</td>
<td>9.00 (8.82)*</td>
<td>3.44 (11.54)</td>
</tr>
<tr>
<td>Convictions</td>
<td>0.11 (0.33)</td>
<td>1.64 (1.74)*</td>
<td>3.25 (3.33)*</td>
<td>1.25 (4.17)</td>
</tr>
</tbody>
</table>

ASPD, antisocial personality disorder; SSPD, schizophrenic-spectrum personality disorder.

* * P < 0.05 vs. SSPD group and v control group; ** P < 0.05 vs. ASPD group in the Bonferroni correction (self-report crime) or the percentile bootstrap method for 20% trimmed means (charges, convictions).
stimuli, although the same bootstrap linear contrast method revealed that the ASPD group demonstrated significantly fewer responses than did controls (P<0.045) during these stimuli. In addition, conventional results indicated that the groups differed significantly in number of responses during the four meaningful stimuli: F(3)=3.442, P=0.021. Post hoc tests (Bonferroni-corrected) indicated that only the SSPD/ASPD group (n=8; mean=0.375, s.d.=0.582) demonstrated significantly fewer responses over the four meaningful stimuli than the control group (n=48; mean=1.990, s.d.=1.435, P=0.018). The bootstrap linear contrast method yielded identical results (P<0.0001).

Amplitude. A repeated-measures MANOVA was used to assess group differences in skin conductance amplitude across the entire orienting phase (group means are displayed in Fig. 2). Results indicated only a trend towards significance: F(3)=2.033, P=0.117. Additional repeated-measures MANOVAs indicated that group differences were non-significant for the first six orienting stimuli (F(3)=1.875, P=0.141) and the four meaningful stimuli (F(3)=1.915, P=0.135).

A boxplot of the skin conductance amplitude data indicated significant skewness and numerous outliers (Fig. 3), which again made appropriate the application of modern statistical methods. A repeated-measures bootstrap method using 20% trimmed means (Wilcox, 2005) was used to assess differences in skin conductance amplitude over the entire orienting phase. Results indicated that the SSPD/ASPD group demonstrated significantly lower amplitude than the control group (P=0.0004). This same repeated-measures bootstrap method was subsequently used to assess differences in skin conductance amplitude over the first six orienting stimuli and over the four meaningful stimuli. Results indicated that the the SSPD group demonstrated significantly lower amplitude over the first six orienting stimuli than controls (P<0.0001), and that the SSPD/ASPD group demonstrated significantly lower amplitude over these stimuli than the ASPD group (P=0.0004), the SSPD group (P<0.0001) and controls (P<0.0001). In addition, results indicated that the SSPD/ASPD group demonstrated significantly lower amplitude over the four meaningful stimuli than controls (P=0.002).

Potential psychiatric confounds

Additional chi-squared analyses were used to assess group differences in comorbid Axis I disorders. Results indicated that groups did not differ significantly regarding the presence of comorbid mood, psychotic or anxiety disorders. Although there was significantly more substance use disorder (i.e. misuse and dependence) among the ASPD and SSPD/ASPD groups (χ²=19.091, P>0.001), further chi-squared analysis revealed that these two groups did not differ significantly in rates of these disorders (χ²=1.257, P=0.262). Consequently, the significantly increased antisocial behaviour and skin conductance orienting to meaningful stimuli found in the comorbid group compared with the ASPD group could not be accounted for by a difference between these groups in substance use disorders.

**DISCUSSION**

This study set out to assess whether there was significant comorbidity between...
schizophrenia-spectrum personality disorders (a group of distinct yet related disorders) and antisocial personality disorder, and if so to determine whether the comorbid group differed in biological (autonomic) and socially meaningful (criminality) ways from individuals with each condition separately. Results indicated not only significant comorbidity but also that the comorbid group was characterised by reduced skin conductance orienting and arousal and more criminal offending. Specifically, the comorbid group demonstrated significantly lower skin conductance orienting response amplitudes to neutral orienting stimuli and reported significantly more criminal behaviour than all of the other groups. Results have potential implications for identifying a distinct subgroup of individuals with schizophrenia-spectrum and antisocial personality disorders, for improving understanding of the reasons for the comorbidity between antisocial behaviour and schizophrenia, and also for the differential treatment, care and management of these individuals in therapeutic and forensic settings.

Findings suggest that ignoring the comorbid link between schizophrenic-spectrum and antisocial personality disorders may obfuscate findings in investigations of either condition separately. Studies of antisocial populations have produced inconsistent results on electrodermal responding, findings that may be clarified when schizophrenia-spectrum disorders are considered as a moderator. For example, it has been observed that although autonomic underresponsivity does not characterise antisociality in general, a specific subgroup of people with schizoid antisocial personality disorder is characterised by autonomic non-response (Rainé & Venables, 1984). Subsequent research has indicated the same pattern of reduced skin conductance orienting in schoolboys with schizotypal-antisocial symptoms and in adults with schizoid psychopathy (Rainé et al., 1999). Conversely, disparate findings of both reduced and increased orienting in people with schizotypal disorder may reflect the need to consider antisociality as a moderating variable, where autonomic underresponding may specifically characterise schizotypal disorder with antisocial tendencies (Rainé et al., 1999). One implication of the current study is that future studies could significantly benefit by assessing both antisocial behaviour and SSPD within the same population in order to elucidate risk factors specific to each of these conditions.

The finding of reduced orienting particularly in the group comorbid for schizophrenia-spectrum and antisocial personality disorder can be viewed within a neuroanatomical context predicated on the frontal cortex. Specifically, skin conductance orienting response is thought to be a marker for structural and functional integrity of the prefrontal cortex, and impaired prefrontal structure and function have been associated with both antisocial behaviour and schizotypal personality. Structural magnetic resonance imaging (MRI) and neurologological studies on humans have demonstrated that reductions in the integrity of the prefrontal cortex (i.e. lesions, or reduced area or volume) are associated with reduced skin conductance responsivity (Critchley, 2002). Functional MRI studies have shown that skin conductance responses during the Iowa gambling task are associated with activation of the ventromedial and orbitofrontal cortex (Critchley et al., 2000). Visual orienting is associated with increased activation in the anterior cingulate, whereas reorienting is associated with increased activation in the middle frontal gyrus (Thiel et al., 2004). Furthermore, stimuli that elicit a skin conductance response, compared with stimuli that do not, result in activation in the hippocampus, anterior cingulate and ventromedial prefrontal cortex (Williams et al., 2000). Although the circuitry underlying the skin conductance response is complex and involves multiple regions, including the right inferior parietal cortex and amygdala, the convergence of findings from structural and functional imaging studies identifies the frontal cortex as a key higher brain area primarily associated with this response.

This linkage of frontal structure and function with the skin conductance response, and the finding in our study that reduced skin conductance orienting response is particularly associated with comorbidity of schizophrenia-spectrum and antisocial personality disorder, suggest the hypothesis that impaired frontal structure and function would be observed in both conditions. There is increasing evidence from both neurocognitive and neuroimaging studies to support this hypothesis. There is growing neuropsychological evidence that juvenile delinquency, antisocial behaviour, criminality and criminal psychopathy are associated with poorer performance on tasks related to both orbitofrontal and ventromedial functioning (Lapierre et al., 1995; Morgan & Lilienfeld, 2000; Brower & Price, 2001; Yechiam et al., 2007). Similarly, poor frontal functioning has been identified as one of the best-replicated neurocognitive correlates of schizotypal personality (Rainé, 2006b). Both structural and functional imaging studies have observed prefrontal impairments in antisocial populations (Rainé, 1997; Rainé et al., 2000) and also in schizotypal populations (Rainé et al., 2002a), although there is stronger evidence for structural impairment in schizophrenia than in schizotypal disorder. Consequently, reduced skin conductance orienting response may particularly characterise the comorbid group because it is a peripheral marker for prefrontal impairment, which in turn represents a common risk factor for both personality disorders. It is proposed that reduced orienting may represent an attentional marker of prefrontal impairment predisposing to both antisocial and schizotypal personality disorders. Taking into account the comorbid relationship between antisocial and schizotypal personality disorders in future studies could help clarify the heterogeneity in findings for both these disorders.

Findings from our study underscore the importance of modern statistical methods such as trimmed means and bootstrapping (Wilcox, 2003), and demonstrate how these techniques may be used to augment conventional analyses when the assumptions of these traditional statistics are violated. In addition, our study illustrates how important group differences may be elucidated using alternative methods when conventional methods fail, and how the nature of data distributions should be considered when choosing an appropriate statistical strategy. Although it is well recognised in the literature that type 1 errors are problematic, it is less well recognised but of importance in the early stages of an enquiry (in this case, understanding comorbidity) that type 2 errors can lead to equally misleading conclusions.

One discrepancy in findings is that although significant results were observed for self-report criminal offending, findings for official crime measures (charges and convictions) were non-significant when comparing the comorbid and ASPD only groups. This may reflect a true result, and may be explained by the fact that official records of charges and convictions are not comprehensive and do not reflect criminal
offences that are undetected. Law enforce-
ment clearance rates clearly indicate that
the substantial majority of criminal activity
goes unsolved (Seigel, 2006). An alternative
interpretation is that there are genuine dif-
f erences between the comorbid and ASPD
only groups, but that the study lacked
power to detect these differences. That this
explanation should not be ruled out is sug-
gested by the fact that the comorbid group
showed a 96.9% increase in criminal
charges and a 98.2% increase in convic-
tions compared with the ASPD only group,
differences that exceed the (statistically sig-
nificant) 56% increase in self-report crime.

If correct, this in turn provides an example
of how – in at least some circumstances –
modern statistical techniques such as
trimmed means (Wilcox, 2005) cannot
entirely compensate for the lack of power
in traditional statistics, although the possi-
bility of genuine null results cannot be

discounted.

Limitations
A limitation of our study may be the rela-
tively small number of stimuli used in the
orienting test. Results from statistical ana-
lyses of only four attentionally meaningful
stimuli should be interpreted with caution.
Clearly, these results need to be replicated,
though they do give provisional findings for
future research. Subsequent studies should
incorporate more stimuli, perhaps expand-
ing upon the use of even more socially
meaningful stimuli such as positively and
negatively charged affective pictures.

One additional consideration to be
made is that the results observed in this
study may merely be attributable to the ad-
effects of both disorders. Although
this is a distinct possibility, the data from
our study may provide some evidence to
the contrary. For example, rates of criminal
offending in the SSPD group were not sig-
ificantly different from those of the con-
tral group (in fact, the former had fewer
charges and convictions than controls);
intuitively the addition of SSPD criminal
offending to that of the ASPD group should
consequently reduce – not increase – levels
of criminal offending. Such results may
speak of a behaviourally (and possibly biologically) distinct subgroup whose dys-
function is more than the additive product of
the two other groups.

Additional research in the area of
comorbidity of schizophrenia-spectrum/
antisocial personality disorder is clearly
warranted. Being the first study of its kind,
this study’s findings require replication.
More efforts are needed to validate the
existence of the comorbidity, especially in
community samples. Future studies should
incorporate the use of corroborating in-
formation (collateral interviews, official
crime records) to enhance the quality of
self-report data.

Implications
Findings may have several implications at
basic research, clinical intervention and for-
ensic levels. At a research level, although re-
searchers have made significant advancements in
the understanding of both schizophrenia-
spectrum and antisocial personality dis-
orders separately, conceptualising those
comorbid for the two as a distinct group
may clarify and strengthen findings for
future investigations not only of this group,
but also of ‘pure’ samples. The importance
of this type of clarification and the explo-
ration of an aetiologically divergent subgroup
with both disorders has been underscored
in the literature (Fowles, 1993). At a clini-
tal level, outcomes of psychophar-
macological treatment programmes may vary markedly between comorbid
and ‘pure’ cases; techniques that demon-
strate effectiveness in both disorders
separately (see Hirose, 2001; Losel, 2001;
Walker et al, 2003; Bilder, 2006) may
be less effective in individuals with co-
morbidity; ultimately, alternative treatment
strategies may need to be developed for this
comorbid group. At a forensic level, com-
morbidity may have practical implications in
the evaluation of dangerousness and poten-
tial for re-offending. If a particular diagnost-
ic entity has been shown to demonstrate
increased rates of criminality, an offender
with such a symptom presentation might
require special sentencing requirements to
both ensure public safety and facilitate
more effective rehabilitation.

In conclusion, results of this initial
study indicate that those with comorbid
schizophrenia-spectrum and antisocial
personality disorders differ in both behav-
ioral and psychophysiological ways from
those with either condition separately, be-
yond the additive effects of both conditions
in combination. The finding of reduced
electrodermal orienting in the comorbid
group confirms and extends findings of
three prior studies observing this same
effect, and may reflect an attentional
resource allocation deficit linked to the
prefrontal cortex which is common to both
clinical groups. Further research on this
comorbid condition is warranted, particularly
because it presents with significantly higher
rates of increased criminal activity.

ACKNOWLEDGEMENTS
This study was supported by Independent Scientist
Award K02 MH01144-01, and grant 5 R03
MH50940-02 from the National Institute of Mental
Health and a grant from the Wacker Foundation to
A.R. We thank Jennifer Babier, Nicole Diamond,
Kevin Ho, Blane Horvath, Sharon Ishikawa, Lori
LaCasse, Todd Lencz, Shari Mills, Kristen Taylor and
Pauline Yaralian for assistance in data collection and
scoring.

REFERENCES
Neuropsychology: A Pocket Handbook for Assessment
(2nd edn) (eds. P. J. Snyder, P. D. Nussbaum &
Association.

frontal lobe dysfunction in violent and criminal
behaviour: a critical review. Journal of Neurology,
Neurosurgery and Psychiatry, 71, 720–726.


Neural activity relating to the generation and
representation of galvanic skin conductance response: a
functional magnetic imaging study. Journal of
Neuroscience, 20, 3033–3040.

Davison, A. C., & Hinkley, D. V. (1997) Bootstrap
Methods and Their Application. Cambridge University
Press.

(1994) Autonomic abnormalities in schizophrenia: state or
trait indicators? Archives of General Psychiatry,
51, 813–824.

Efron, B. & Tibshirani, R. J. (1993) An Introduction to
the Bootstrap. Chapman & Hall.

Research Institute.

First, M. B., Gibbon, M., Spitzer, R. L., et al (1979a)
User’s Guide for the Structured Clinical Interview for DSM–
IV Axis I Disorders (SCID–I). American
Psychiatric Press.

User’s Guide for the Structured Clinical Interview for DSM–
IV Axis I Disorders – Clinician Version (SCID–CI).
American Psychiatric Press.

Fowles, D. C. (1995) Electrodermal activity and
antisocial behavior: empirical findings and theoretical
issues. In Progress in Electrodermal Research (eds C. Roy,

Hirose, S. (2001) Effective treatment of aggression and
impulsivity in antisocial personality disorder with
risperidone. Psychiatry and Clinical Neurosciences, 55,
161–162.

Hollingshead, A. B. (1975) Four Factor Index of Social
Status. Department of Sociology, Yale University


Impact and clinical significance of a preventive intervention for disruptive boys
15-year follow-up

RACHEL BOISJOLI, FRANK VITARO, ÉRIC LACOURSE, EDWARD D. BARKER and RICHARD E. TREMBLAY

Background Many intervention programmes have attempted to reduce disruptive behaviour problems during early childhood to prevent maladjustment during adolescence and adulthood.

Aims To assess the long-term impact and clinical significance of a 2-year multicomponent preventive intervention on criminal behaviour and academic achievement, using intention-to-treat analyses.

Method Targeted disruptive – aggressive boys considered to be at risk of later criminality and low school achievement (n=250), identified from a community sample (n=895), were randomly allocated to an intervention or a control group. The rest of the sample (n=645) served as the low-risk group. The intervention was multimodal and aimed at boys, parents and teachers. Official data measured both outcomes.

Results Significantly more boys in the intervention group (13%; P<0.05) completed high-school graduation and generally fewer (11%; P=0.06) had a criminal record compared with those allocated to the control group.

Conclusions The results suggest that early preventive intervention for those at high risk of antisocial behaviour is likely to benefit both the individuals concerned and society.

Declaration of interest None.

Disruptive behaviour problems (aggressive, hyperactive and oppositional behaviours) during early childhood predict maladjustment (e.g. violence, criminality and school drop-out) during adolescence and adulthood (Farrington, 1992; Tremblay et al, 1992a; Moffitt, 1993; Patterson, 1996; Ferguson et al, 2002; Loeb, 2002; Vitato et al, 2005). Intervention programmes have attempted to reduce disruptive behaviours to prevent such negative outcomes (e.g. Durlak & Wells, 1997; Farrington & Welsh, 2003). However, their long-term effects are rarely evaluated. Moreover, a significant long-term effect is not sufficient for claiming the efficacy of an intervention; its clinical significance has to be assessed by comparing its effects with a normative or low-risk group.

This study’s main objective was to evaluate whether participation in a preventive intervention targeting early disruptiveness predicted a higher rate of high-school graduation and a lower rate of crime involvement compared with the control group, by age 24 years. The second objective was to verify whether the boys who received the intervention would resemble the boys in the low-risk group with regard to the outcomes, whereas the boys in the control group would not.

METHOD

The global objective of the Montreal Longitudinal Experimental Study (Tremblay et al, 1992b) was to examine prospectively the development of a large sample of boys attending inner-city kindergartens who had backgrounds of low socio-economic status, with a particular focus on antisocial behaviour and school adjustment. Behaviour ratings of male pupils, mean age 6.1 years (s.d.=0.32), were obtained from 87% of the kindergarten teachers in 53 schools in areas of low socio-economic status in Montreal, Canada, at the end of the 1984 school year. A total of 1161 boys were rated. After exclusion of pupils who did not meet additional selection criteria – i.e. ethnicity (only boys with Canadian-born parents whose first language was French were included) and education (only boys whose parents had 14 years or less of schooling were included) – that number was reduced to 895. The purpose of these additional selection criteria was to create a homogeneous sample (through methodological control).

Boys were assessed by their kindergarten teacher by means of the Social Behavior Questionnaire (SBQ; Tremblay et al, 1991). This contains 38 items grouped into four components: disruptive (13 items), anxious (5 items), inattentive (4 items) and prosocial (10 items). The disruptiveness scale (x=0.93) includes three categories of behaviour (Loeb et al, 1989): aggression (3 items), oppositional behaviour (5 items) and hyperactivity (2 items), and was used to identify at-risk children. From the total sample, boys with a score above the 70th percentile (n=250) on the disruptiveness scale were considered to be at risk of later antisocial behaviour and dropping out of school (White et al, 1990; Tremblay et al, 1992a). Although this cut-off point is somewhat arbitrary, it has been used successfully to predict serious maladjustment in this sample (Tremblay et al, 1994). These 250 boys were randomly assigned to one of three groups (prevention, n=69; attention-control, n=123; control, n=58) by drawing the names from a box until the necessary numbers were obtained. Given that no difference was found between the two control groups on any outcome during adolescence or early adulthood (see below), they were combined into a single control group for later analyses (Fig. 1). The attention-control group was equivalent to a no-treatment sensitisation or contact control group; the control group was a no-treatment, no-contact control group.

Among these, 172 families (69%) agreed to participate in the intervention programme, but all the at-risk boys (n=250) were kept in the longitudinal study and their data were included in the intention-to-treat analyses. Both the boys and their families participated in the intervention programme. The rest of the larger sample, representing participants who obtained scores below the 70th percentile (n=645), were considered to be at lower risk and were kept in the study to test the clinical significance of the prevention programme.
Preventive intervention programme

Three foci of the applied preventive intervention programme were based on a literature review addressing early intervention with aggressive children before 1984. The first theme identified was social skills training for the disruptive boys (Cartledge & Milburn, 1980; Kettlewell & Kausch, 1983; Michelson et al., 1983; Schneider & Byrne, 1987). Social skills training aimed at promoting changes in behaviour towards peers, yielding more social acceptance and less inclination towards antisocial peers. Training was offered at school in small groups of four to seven children, with a ratio of three prosocial children from the school to one disruptive child in each group. The second focus was that of parent training in effective child-rearing, based on the Oregon Social Learning Center Model (Patterson et al., 1975). The third domain was the provision of information and support for teachers concerning at-risk boys, which served as a complement to the parent training.

The intervention programme lasted 2 school years, from September 1985 to June 1987. Boys were 7 years old when the intervention started and 9 years old when it ended. A detailed description of the treatment is presented elsewhere (Tremblay et al., 1992b).

Implementation assessment

In order to evaluate programme exposure, the therapist responsible for each child–family–teacher unit indicated at the end of each planned training session whether or not the session had taken place and the percentage of content that had been delivered in the session with regard to the pre-planned, standardised content. Over 85% of the children who participated in the intervention attended at least two-thirds of the social skills training sessions. The maximum number of sessions given to the parents was 46, with the mean number of sessions for the duration of the programme being 17.4, including parents who discontinued their participation in the programme. Parents were given as many sessions as needed to master the skills, following the adaptive preventive intervention approach proposed by Collins et al. (2004). However, 75% of the parents covered at least two-thirds of the content and objectives of the planned training programme. Teachers demonstrated low interest and limited availability; they were generally not able to spend much time discussing teaching strategies for one child. Therefore, meetings with teachers were fewer than planned (about 50% of teachers participated in at least one meeting). Work with the parents and teachers was carried out by full-time trained therapists: two university-trained childcare workers, one psychologist and one social worker. Social skills training sessions were taped and used for weekly feedback and to maintain the integrity of the programme across therapists.

Control and outcome measures

Control variables assessed in kindergarten

Although no significant difference was found between the intervention group and the control group after random assignment, two control variables – parental occupational prestige and children’s disruptiveness – were included in the analyses to completely level initial differences and reduce bias in estimating the impact of the intervention programme. Parental prestige was established using fathers’ and mothers’ occupational status at pre-test and used as an indicator of family background. It was calculated using the Canadian socio-economic status index of Blishen et al. (1987). This variable is known to be linked to behaviour problems and delinquency and to high-school graduation (Huesmann et al., 1984). The children’s disruptiveness variable used for selection and pre-test was also used as a control variable.

Outcome measures collected at age 24 years

A high-school diploma was selected as the measure of scholarly achievement. This variable was used instead of school dropout or non-age-appropriate regular classroom placement, previously used to assess school performance (Vitaro et al., 2001), because it represents a more definite measure; some boys who dropped out of high school returned to complete their education and received a diploma. The Ministry of Education of Quebec confirmed the award of a high-school diploma as of year 2003 for 879 persons in the original sample, including 242 of the original 250 participants in the prevention or control groups. This categorical variable provided information on whether or not the participants had obtained a high-school diploma by age 24 years. Overall, 427 of the 879 participants (48.6%) had done so.
Possession of a criminal record was selected as the measure of crime involvement. Criminal records were obtained from official files as of year 2003 for all of the 895 persons in the original sample, including the 250 participants in the prevention or control groups. This categorical variable provided information on whether or not the participant had a criminal record by age 24 years. Of the 895 participants, 178 (19.9%) had acquired a criminal record by age 24 years. Criminal offences were divided into five categories, as defined by the Ministry of Public Security of the Province of Quebec (prevalence for each category is shown in parentheses): crimes against persons, e.g. homicide (17.9%); property crimes, e.g. arson (31.2%); other Criminal Code offences, e.g. prostitution (25.5%); motor vehicle-related offences, e.g. impaired driving (8.8%); and drugs and narcotics-related offences, e.g. possession (16.4%).

Analyses
Two sets of analyses were performed, after verifying that the data did not violate any of the assumptions of logistic regression. For the first set of logistic regressions, achieving a high-school diploma and presence of a criminal record were separately regressed on the experimental conditions (i.e. intervention vs. control) while controlling for parental occupational status and disruptiveness. For the second set of logistic regressions, the same outcomes were regressed on group membership (i.e. intervention and control groups, plus the low-risk group) while controlling for parental occupational status. In order to test the effectiveness of the programme, all participants in the intervention sample were included in the intention-to-treat analytic strategy, whether they received the intervention or not.

RESULTS

Differences between control and intervention groups

Frequencies of high-school graduation and criminal records are presented in Table 1.

High-school graduation

After controlling for parental occupational status and initial level of children’s disruptiveness, we found that being in the intervention group was associated with a higher rate of high-school graduation than being in the control group ($\beta = 0.78$, OR = 2.19, Wald $\chi^2 = 6.06; P < 0.05$).

Table 1  

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Intervention group</th>
<th>Normative group</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-school graduation¹</td>
<td>56 (32.2)</td>
<td>31 (45.6)</td>
<td>340 (53.4)</td>
</tr>
<tr>
<td>Criminal record²</td>
<td>59 (32.6)</td>
<td>15 (21.7)</td>
<td>104 (16.1)</td>
</tr>
</tbody>
</table>

1. Includes all available data for the original sample (n = 879).  
2. Includes all available data for the original sample (n = 895).

Criminal record

Being in the intervention group was marginally associated with a lower rate of criminal record than being in the control group ($\beta = -0.65$, OR = 0.52, Wald $\chi^2 = 3.68; P = 0.06$).

Differences between experimental groups and the low-risk group

High-school graduation

After controlling for parental occupational status, being in the intervention group compared with being in the low-risk group predicted a similar rate of high-school graduation ($\beta = -0.19$, OR = 0.83, Wald $\chi^2 = 0.52$; NS), but being in the control group compared with being in the low-risk group predicted a lower rate of high-school graduation ($\beta = -0.84$, OR = 0.43, Wald $\chi^2 = 20.77; P < 0.0001$).

Criminal record

Being in the intervention group compared with being in the low-risk group predicted a similar rate of criminal record ($\beta = 0.30$, OR = 1.35, Wald $\chi^2 = 0.92$; NS), whereas being in the control group compared with being in the low-risk group predicted a higher rate of criminal record ($\beta = 0.89$, OR = 2.45, Wald $\chi^2 = 21.69; P < 0.0001$).

DISCUSSION

The first goal of our study was to use intention-to-treat analyses to assess the long-term impact of a multicomponent preventive intervention programme targeting disruptive boys from homes with low socio-economic status; these boys were considered at high risk of low academic achievement and chronic antisocial behaviour. The impact of the programme was evaluated by contrasting disruptive boys who participated in the preventive programme and their counterparts in a control group on two outcomes: high-school graduation by age 24 years and official criminal records. The second goal was to compare the intervention group and the control group with the rest of the boys in the low socio-economic status group who initially scored below the 70th percentile on disruptiveness (i.e. the low-risk group).

Impact and social significance of the programme

As predicted, a positive effect of the intervention programme was found for high-school graduation. The likelihood of having a high-school diploma was more than twice as high for the intervention group as for the control group. These results support earlier findings during adolescence on school drop-out (Vitaro et al., 1999). Although marginal, a positive effect of the intervention was also found for possession of a criminal record: the likelihood of having a criminal record was almost twice as high for the control group as for the intervention group.

Comparing the experimental groups with a low-risk group on high-school graduation allows evaluation of the clinical significance of the intervention. In addition to the significant effect of the intervention on high-school graduation when compared with the control group, being in the intervention group predicted a rate of high-school graduation similar to that of the low-risk group. In the same way, the intervention group obtained a similar rate of criminal record as the low-risk group, whereas the risk of having a criminal record in the control group was more than double that for the low-risk group. These results confirm the relevance of reducing early disruptiveness to prevent later adjustment problems, and highlights the predictive power of early disruptiveness in an experimental clinical context.

Considering that, in adolescence, a significantly greater percentage of boys in the prevention group remained in an age-appropriate regular classroom compared with controls (Vitaro et al., 1999), and that the level of delinquency was higher for the control group compared with the intervention group on high-school graduation and possession of a criminal record.
group (Lacourse et al., 2002), these results are not surprising. However, although encouraging, these findings should be considered in light of the fact that the rate of high-school graduation for the intervention group was only 46%, and the rate of having a criminal record was as high as 22%. In comparison, the rate of high-school graduation in the low-risk group was also low (68%) and the rate for criminal record was also high (16%), bringing the rates for the whole sample to 49% for high-school graduation and 19% for possessing a criminal record. In consequence, although boys in the intervention group became similar to their low-risk peers with respect to high-school graduation and criminal activities, the burden of other risk factors (i.e. low socio-economic status, inner-city residence) took its toll on the whole sample. It is thus important to acknowledge that a preventive intervention programme, albeit intensive, multimodal and long-term, has only a limited protective effect under the conditions of chronic socio-familial adversity and environmental risk.

**Limitations**

A number of limitations have to be considered. First, this study used only one measure of antisocial behaviour. Official records used in this study can be considered as a good indicator of antisocial behaviour, but their interpretation is limited since they provide no direct information on observable behaviours. On the other hand, the use of this measure resulted in low attrition. It is also convenient for cost-effectiveness and clinical significance analyses. Second, the sample was restricted to French-speaking male participants of low socio-economic status. Generalisability is therefore limited. A similar intervention with a mixed sample from a middle-class environment could generate different results and yield different conclusions. Finally, potential moderators and mediators still have to be explored.

**Implications of the study**

Despite these limitations, our study contributes to the critical need for long-term follow-up investigations by giving a valuable and rare picture of the long-term effects of an early preventive programme. This research also allowed the clinical significance of the programme to be tested by comparing the intervention and the control groups with a group of peers from the same high-risk environment. Given the cost to society of criminality and failure to graduate from high school (Kerrick & Bell, 1998), this study also stresses the cost-effectiveness of preventive intervention even if no formal examination of cost-effectiveness was performed.

Taking into account these results, some considerations can be put forward. As suggested by Tremblay et al. (1996), a longer intervention or a booster programme covering the transition to high school and into adulthood might have resulted in more robust effects during adulthood. In other words, the duration of the intervention (2 years) may not be sufficient or optimal, particularly when the external conditions are unfavourable. Several authors (Reid, 1993; Lochman & Wells, 1996) have suggested that an intervention should last for at least the whole elementary schooling period. As for the number of components, most experts agree on the importance of targeting different systems in children’s life, such as parents, teachers and the children themselves, as in the present study. However, additional systems such as peer groups should be targeted in future studies, in order to modify the additional important sources of influence that affect the development of antisocial behaviour (Coie & Jacobs, 1993; Greenberg et al., 2001; Boivin et al., 2005). Improving external conditions would also represent a good course of action for improving the impact of a child, family and school-centred preventive intervention.

**ACKNOWLEDGEMENTS**

This research was supported by the Social Sciences and Humanities Research Council of Canada (grant 752-2005-1882).

**REFERENCES**


Structured patient–clinician communication and 1-year outcome in community mental healthcare

Cluster randomised controlled trial

STEFAN PRIEBE, ROSEMARIE McCabe, JENS BULLENKAMP, LARS HANSSON, CHRISTOPH LAUBER, RAFAEL MARTINEZ-LEAL, WULF ROSSLER, HANS SALIZE, BENGT SVENSSON, FRANCISCO TORRES-GONZALES, ROB VAN DEN BRINK, DURK WIERBOS and DONNA J. WRIGHT

Background Patient–clinician communication is central to mental healthcare but neglected in research.

Aims To test a new computer-mediated intervention structuring patient–clinician dialogue (DIALOG) focusing on patients’ quality of life and needs for care.

Method In a cluster randomised controlled trial, 134 keyworkers in six countries were allocated to DIALOG or treatment as usual; 507 people with schizophrenia or related disorders were included. Every 2 months for 1 year, clinicians asked patients to rate satisfaction with quality of life and treatment, and request additional or different support. Responses were fed back immediately in screen displays, compared with previous ratings and discussed. Primary outcome was subjective quality of life, and secondary outcomes were unmet needs and treatment satisfaction.

Results Of 507 patients, 56 were lost to follow-up and 451 were included in intention-to-treat analyses. Patients receiving the DIALOG intervention had better subjective quality of life, fewer unmet needs and higher treatment satisfaction after 12 months.

Conclusions Structuring patient–clinician dialogue to focus on patients’ views positively influenced quality of life, needs for care and treatment satisfaction.

Declarations of interest None.

Funding detailed in Acknowledgements.

Regular meetings between a patient and their clinician are at the heart of community mental healthcare. They are used to communicate the patient’s condition, personal situation and ongoing treatment. This routine communication has been subjected to little systematic research, and there is no evidence-based method to structure the communication to enhance long-term treatment outcome. A simple communication checklist completed by patients before seeing their clinician led to improved communication and treatment changes, but its effect on long-term outcome was not assessed (Van Os et al, 2004). New technologies to support patient–clinician communication are increasingly used in healthcare, although rarely in community psychiatry (Ahmed & Boivert, 2006). Hence, we devised a computer-mediated procedure to structure patient–clinician dialogue (DIALOG) to ensure that a range of life domains and treatment aspects were consistently addressed and patients’ perspectives always elicited. The results were fed back immediately and were intended to feed directly into patient–clinician discussions and shape subsequent care.

METHOD

The aim of this study was to investigate whether using the new intervention regularly in routine meetings between clinicians and patients with schizophrenia in the community would be associated with more favourable quality of life, fewer unmet needs for care and higher treatment satisfaction after a 1 year period compared with treatment as usual. The hypothesis was tested in a cluster randomised trial in six European countries (trial number ISRCTN75571732).

Settings This study was conducted in community psychiatric services in Granada (Spain), Groningen (The Netherlands), London (UK), Lund (Sweden), Mannheim (Germany) and Zurich (Switzerland) covering urban and mixed urban–rural areas. The number of participating teams per country varied between two (Lund) and six (London).

All teams were multidisciplinary and provided comprehensive care programmes for people with severe and enduring mental illnesses. They operated a keyworker system in which every patient has a designated clinician working within a team but with lead responsibility for care coordination and delivery. Referrals were determined by residency in the catchment area and age (18–63 years).

Participants Eligibility criteria for participating clinicians were a professional qualification in mental health or a minimum of 1 year’s professional experience in an out-patient setting, and an active case-load as keyworker. The case-loads of participating clinicians were screened to identify suitable patients meeting the following inclusion criteria: living in the community (not 24 h supported accommodation) and treated as out-patients by community psychiatric teams; at least 3 months of continuous care in the current service; capable of giving informed consent; having sufficient knowledge of the language of the host country; having a primary diagnosis of schizophrenia or related psychotic disorder (ICD–10 F20–F29); aged between 18 and 65 years; having routinely at least one meeting with their keyworker every 2 months with the expectation that they would continue with the service for the next 12 months; and having no severe organic psychiatric illness or primary substance misuse. Patients were first informed about the study by clinicians and then – if they agreed – approached by a researcher for consent. The study was approved by relevant ethics committees in the six countries, and written informed consent was obtained from all clinicians and patients.

Design and process of randomisation

The intervention was evaluated using a cluster randomised controlled trial design. Clinicians were randomly assigned to either the intervention or treatment as usual, with a pre-post design over a 1-year period.
Cluster randomisation was used to avoid potential contamination between the interventions in the two groups. Clinicians were randomised by computer-generated random block number allocation sequence to ensure an equal balance across sites. The randomisation procedure was completed separately for each country and team. A researcher not involved in the study generated the random allocation sequence. The process of allocating clinicians to the treatment as usual or intervention groups was by numbered, sealed envelopes. Masking of researchers to the allocation of the patients was attempted for the duration of the study. As masking was expected to be difficult to maintain, interviewers’ awareness of patients’ allocation was documented and assessed at the end of the study. In four countries all eligible patients from participating clinicians were asked to take part in the study. In the remaining two countries where clinicians had considerably higher patient case-loads, a maximum random sample of 12 patients was taken per clinician.

**Intervention**

Clinicians in the control group continued with standard treatment with their participating patients. Clinicians in the intervention group, in addition to continuing with standard treatment with their participating patients, also implemented the new manualsed intervention. In the intervention group clinicians used DIALOG, a computer-mediated procedure to discuss 11 domains with their patients. They asked patients to rate their satisfaction with eight life domains (mental health, physical health, accommodation, job situation, leisure activities, friendships, relationship with family/partner, personal safety) and three treatment domains (practical help, psychological help and medication). Each satisfaction item was rated on a rating scale of 1–7, from ‘couldn’t be worse’ to ‘couldn’t be better’, and followed by a question on whether the patient wanted any additional or different help in the given domain. If the patient answered yes, the type of the requested additional or different support was recorded. The 11 domains were presented in a fixed order and an explicit response was required for each item before proceeding to the next item.

Patients’ answers to all questions were entered directly onto a hand-held computer or laptop using software specifically developed for the study over a 2-year period.

Figure 1 illustrates possible screen displays, taking accommodation as an example (all of the other 10 domains can be displayed in the same way). A single domain could be viewed with the current rating compared with the rating 2 months previously. The domain could be viewed in the context of all the other domains in a summary graph comparing previous and current ratings for all 11 domains (end of Fig. 1). All 11 domains could also be viewed as a list in a summary table showing number of points change since the last meeting (e.g. +2, −3).

The intervention was applied every 2 months in meetings that had been arranged as part of routine care. The new procedure was designed to alter interactions so that the patient’s views on their situation and needs for care were the central point of treatment discussions and the patient’s view on what kind of help would improve their current situation was made explicit. Patients and clinicians discussed current and previous ratings, reasons for change and what kind of additional or different support might be helpful. The underlying rationale was that providing patients and clinicians with this information would lead to explicit negotiation about what the patient wanted and what the clinician could do about it. This, in turn, would improve subsequent care and the patient’s quality of life.
Each clinician in the intervention group was individually trained to use the software by a researcher and provided with written instructions. They were instructed on how the ratings should be used to facilitate a dialogue with the patients, particularly when there were changes since the last rating, explicit dissatisfaction with life domains or treatment aspects, or the patient wanted additional or different support.

Data collection
Collection of baseline data began in December 2002 and post-intervention data collection ended in May 2005. At both time points clinicians and patients were interviewed by researchers who had no involvement in the patients’ care. Patients were interviewed either at the team office or at home, according to their preference.

Outcomes
Outcome in the two groups was compared in a pre–post design. Primary outcome was subjective quality of life (SQOL) at 12 months controlling for baseline score. Quality of life was measured using the Manchester Short Assessment of Quality of Life (MANSA; Priebe et al, 1999) whereby patients rate their satisfaction with life in general and different life domains on Likert-type scales ranging from 1 (couldn’t be worse) to 7 (couldn’t be better), an approach that is consistent with the Quality of Life Interview (Lehman, 1988). The mean score of all 12 satisfaction ratings is taken as the indicator of SQOL.

Secondary outcomes were the number of unmet needs for care and satisfaction with treatment at 12 months, controlling in each case for the baseline score. Need for care was measured on the Camberwell Assessment of Need Short Appraisal Schedule, patient-rated version (CANSAS; Slade et al, 1996) which assesses health and social needs across 22 domains. For each domain it distinguishes between ‘no need’ (rating of 0), ‘met need’ (rating of 1) and ‘unmet need’ (rating of 2). Patients’ satisfaction with treatment was assessed on the Client Satisfaction Questionnaire (CSQ–8; Nguyen et al, 1983), which consists of eight items rated from 1 to 4 (with higher scores indicating greater treatment satisfaction).

Interviewers assessed patients’ symptoms on the 30-item Positive and Negative Syndrome Scale (PANSS; Kay et al, 1987). The scale assesses positive, negative and general symptoms and is rated on a scale of 1–7 (with higher scores indicating more severe symptoms). Socio-demographic and clinical characteristics of patients were obtained at baseline. Psychiatric diagnosis was obtained through a standardised and computer-based method using operationalised criteria (OPCRIT; McGuffin et al, 1991). Researchers received training in all rating procedures and achieved good inter-rater reliability using videotaped interviews for PANSS (Cohen’s kappa 0.71) and case vignettes for CANSAS (0.90).

Statistical analysis
R version 2.2.0 (Ihaka & Gentleman, 1996) was used to compare the intervention and control groups in an intention-to-treat analysis. Descriptive statistics are presented, with frequency and percentage distributions for categorical data and means and standard deviations for continuous data.

In the main analyses patients were excluded only if they gave no information at follow-up. A sensitivity analysis using multiple imputation was also carried out to check the effect of excluding these patients. Each outcome was analysed using a mixed-effects model with baseline score for that variable, treatment allocation and length of follow-up as fixed effects, and centre and keyworker as random effects. Length of follow-up was considered as a potentially confounding covariate that might have introduced post-randomisation variance, and centre and keyworker were included in the model to adjust for the effect of clustering. Results are presented as 95% confidence intervals. Assumptions were checked graphically. Effects in the linear mixed-effects model are reported as partial eta squared, which is the proportion of total variability attributable to a factor.

Sample size
We aimed to obtain complete data for 240 patients in each group. With a significance level of $\alpha=0.05$, this sample size would allow the detection of an effect size of 0.2 with 59% power, and of an effect size of 0.5 with more than 99% power.

RESULTS
Participant flow
One hundred and thirty-four clinicians consented to take part in the study, of whom 64 were randomised to the intervention group and 70 to the control group. From their case-loads, 507 eligible patients agreed to take part, with 236 patients in the treatment as usual and 271 in the intervention group. The number of patients per clinician ranged from 1 to 12 (mean 3.73).

At 12 months, 451 patients (243 intervention, 208 treatment as usual) were re-interviewed (88.9% follow-up). There were 17 keyworker changes during the study, with only one replacement clinician not agreeing to participate. Patient flow during the trial is shown in Fig. 2.

The baseline to follow-up period ranged between 8 and 20 months (mean 12.4, s.d.=1.68 months). The range reflects late recruitment (16 patients had a follow-up of less than 12 months) and difficulties contacting patients and arranging follow-up interviews. For 283 (62.7%) out of the 451 re-interviewed patients, researchers stated they knew their allocation, making the correct assumption in 275 cases. Masking had been compromised through information that was revealed in previous contacts of researchers with the teams or in their assessments of the patients.

The mean number of meetings with structured communication in the intervention group was 5.21. Four patients had no such meeting, 12 patients had one, 14 had two, 15 had three, 40 had four, 45 had five, 46 six, and 95 had seven meetings. The time of all meetings between keyworkers and patients was documented over a 2-month period (i.e. months 6 and 7 of the 12-month study period), and the total time spent by keyworkers in meetings with each other showed no significant difference between the two groups (intervention group, mean 240, s.d.=201.9 min; control group, mean 251, s.d.=199.2 min).

An intention-to-treat analysis was conducted with the analysis set including all patients with at least one post-randomisation observation.

Baseline characteristics of participants
The characteristics, both socio-demographic and clinical, of clinicians and patients are shown in Table 1. There were no significant differences in the characteristics of participants in the control and intervention groups.

Outcomes
Outcomes are summarised in Table 2. At 12-month follow-up patients in the intervention group had significantly higher
Structured Patient–Clinician Communication and 1-Year Outcome

SQOL scores, fewer unmet needs and higher treatment satisfaction compared with patients in the control group. The effect sizes based on adjusted means and standard deviations for the three outcomes vary between 0.20 and 0.27. 

Owing to the floor effect for unmet needs and ceiling effect for quality of life, a substantial improvement was unlikely to be achieved in those patients who already had a positive SQOL and few unmet needs at the beginning of the trial. We therefore conducted a post hoc analysis on the group as a whole, with those patients who at baseline had at least two unmet needs and a SQOL score lower than 5 (i.e. ‘mixed’ or lower). In those 195 patients (106 in the intervention and 89 in the control group), the effect size in relation to SQOL was 0.43 (adjusted mean difference 0.33, \( P = 0.006 \)) and in relation to unmet needs was 0.52 (adjusted mean difference 1.16, \( P = 0.003 \)). 

As a sensitivity analysis we fitted the same models imputing the missing outcomes using regression, using five sets of imputations. The resulting effect sizes were almost unchanged. The two groups showed no statistically significant difference in any of the psychopathology scores on the PANSS.

DISCUSSION

This study tested the effectiveness of a novel intervention in community care of patients with schizophrenia and related psychotic disorders. This is the first study to change the structure of patient–clinician interaction in community mental healthcare across a range of healthcare systems and to test its effect on long-term outcomes of care. After 12 months, the intervention had a significant positive effect on all three outcomes (i.e. quality of life, unmet needs for care and treatment satisfaction). Previous studies that structured communication between patients and clinicians were based on only a few patients (Ahmed & Boisvert, 2006) or did not assess its effect on long-term outcome of care (Van Os et al., 2004). This study using a large sample across different healthcare systems demonstrated the efficacy of computer-mediated communication on outcome over a 1-year period.

This intervention ensured that 11 life and treatment domains were consistently addressed and patients’ views and priorities

Fig. 2 Trial CONSORT diagram. In two centres a maximum random sample of 12 patients was taken per clinician owing to a high patient case-load.
were always considered (Rosenheck et al., 2005). This is likely to have increased awareness of patients’ views and their changes over time, resulting in care that reduces unmet needs and increases SQOL and treatment satisfaction (Lasalvia et al., 2005). This was achieved although symptom levels did not change. Given the enduring nature of the disorders in our sample, this was as expected and suggests that patients’ quality of life can be improved even when symptoms do not show significant change (Holloway & Carson, 1998; Trieman et al., 1999).

### Limitations and strengths

The study should be considered in the light of its limitations. Participating teams and clinicians might not have been representative of the given mental healthcare systems. The novel intervention was not consistently administered, as evidenced by the variation in the number of structured communications for individual patients (although with a mean of approximately 5 per patient), which reflects the pragmatic nature of the trial. Finally, masking of interviewers could not be maintained for the majority of patients, and exclusively subjective measures were used as outcome criteria.

The strengths of the study are that the intervention was tested under routine conditions and in six European healthcare settings, with high follow-up rates of 90% in this often difficult to reach and mobile population. The intervention requires little additional investment and minimal training of clinicians. It did not significantly increase the time spent by keyworkers and patients in meetings with each other, and was viewed favourably by both patients and keyworkers (see online supplement to this paper). It can be applied without reconfiguration of services and would be easy to implement widely. We found a positive effect in a sample with predominantly long-term problems – the mean length of illness was more than 15 years – and the scope to achieve substantial improvements of SQOL in such samples over a 1-year period is usually regarded as somewhat limited.

### Intervening in patient–clinician communication

So far, there is a paucity of evidence-based interventions that can be used in routine meetings between clinicians and people with schizophrenia to enhance quality of life (Marshall et al., 2004; Slade et al., 2006). The intervention tested in this study targets patient–clinician communication as the central component of care delivery and structures it in a patient-centred manner. There is evidence that the quality of patient–clinician communication plays a role in treatment outcome. In primary care consultations, a positive patient-centred approach was associated with higher patient satisfaction, less symptom burden and fewer referrals to other services (Little et al., 2001). In mental healthcare, a simple communication checklist completed by patients before seeing their doctor, where they indicated which of 20 common needs they wanted to discuss, led to improved patient–doctor communication and changes in treatment (Van Os et al., 2004).

The use of computers was also found to facilitate communication between clinicians and people with schizophrenia. Specifically,
patients’ responses to structured questions concerning treatment goals and expectations were visually presented and reviewed on a computer screen. This improved discussion of treatment and the identification of realistic goals for therapy (Ahmed & Boisvert, 2006). The authors proposed that using both visual and auditory techniques may facilitate communication by improving patient attention, information assimilation and reducing interference from psychiatric symptoms such as delusions.

The current intervention is simple, non-intrusive and inexpensive. Although the effect sizes in this study were small, they do not indicate a dramatic change in the treatment outcome. The effect sizes of 0.43 and 0.52 were achieved through the DIALOG intervention. These do not indicate a dramatic change in the living situation of patients on a group level but suggest a real difference for at least some of the patients. It remains unclear to what extent this effect is due to: (a) the mere structuring of the meeting which ensures that important areas are always covered; (b) the focus on patient views of outcome in the meeting; and (c) the specific computer-mediated option of comparing current ratings with previous ratings across different life domains.

If used in routine settings the intervention might facilitate the generation of regular outcome data. As the procedure involves the assessment of central outcome criteria in community psychiatry (i.e. satisfaction with life domains and with treatment), these scores may feed into processes of quality management and service evaluation (McCabe & Priebe, 2002; Priebe et al., 2002). Gathering outcome data from a procedure that is meaningful to patients and clinicians and beneficial for the individual patient is more likely to be successful than conventional methods of routine outcome measurement in which outcomes are rated by patients outside clinical consultations and the results later made available to clinicians (Gilbody et al., 2001; Slade et al., 2006). The latter approach makes it difficult to determine whether the process of outcomes management had an impact on what clinicians and patients did in clinical consultations. Incorporating the assessment and feedback of outcomes into routine clinical encounters makes it more likely to have a direct impact on what happens in practice when clinicians and patients interact.

In conclusion, a simple computer-mediated procedure to structure routine communication between patient and clinician can have a significant positive effect on treatment outcome over a 1-year period in patients with schizophrenia in the community. Future studies should test the feasibility and effectiveness of similar procedures for improving patient–clinician communication with other patient groups and in other out-patient settings. Moreover, qualitative and experimental research might help to develop interventions that are more effective than DIALOG in influencing both the therapeutic communication and outcome, and identify the mediating processes between better communication and more favourable outcome.

ACKNOWLEDGEMENTS

The Towards More Effective European Community Care for Patients with Severe Psychosis (MECCA) group includes: Marta Ribes Leyva, Maria F. Soriano Peña, Beatriz Arroyo de Domingo (Granada); Kerstin Wolters, Aukeliain Mulder, Jappie Tiersma (Groningen); Rakhee Haque (London); Tommy Björkman (Lund); Marita Reichenbacher, Anette Axén (Mannheim); Patric Meyer, Minka Burgi (Zurich). We thank the clinicians and patients who took part in the Mediating Community Care study, Dr. Michael Dewey for statistical advice and Dr. Patrick Healey and Greg Mills for software development and technical support. The study was funded by the Research Directorate of the European Commission within the Framework Programme 5 (QLG5-CT-2002-01938).

REFERENCES


Table 2 Differences in quality of life, treatment satisfaction and unmet needs between groups at 12-month follow-up

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Treatment as usual</th>
<th>Linear mixed-effects model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted mean (s.e.)</td>
<td>Adjusted mean (s.e.)</td>
</tr>
<tr>
<td>MANSAn scores</td>
<td>240 4.86 (0.04) 208 4.74 (0.04)</td>
<td>0.12 (0.00 to 0.26)</td>
</tr>
<tr>
<td>CSQn scores</td>
<td>241 4.87 (0.87) 208 4.72 (0.88)</td>
<td>0.92 (0.22 to 1.56)</td>
</tr>
<tr>
<td>CANSASn scores</td>
<td>241 25.96 (0.26) 207 25.04 (0.27)</td>
<td>0.41 (−0.79 to −0.01)</td>
</tr>
<tr>
<td>PANSS sub-scale scores</td>
<td>241 2.05 (0.15) 208 2.46 (0.16)</td>
<td>−0.46 (−2.19 to 1.27)</td>
</tr>
</tbody>
</table>

MANSa, Manchester Short Assessment of Quality of Life; CSQ, Client Satisfaction Questionnaire; CANSAS, Camberwell Assessment of Need Short Appraisal Schedule, patient-rated version; PANSS, Positive and Negative Syndrome Scale.


Clinical effectiveness of treatments for anorexia nervosa in adolescents

Randomised controlled trial†

SIMON G. GOWERS, ANDREW CLARK, CHRIS ROBERTS, ALISON GRIFFITHS, VANESSA EDWARDS, CLAUDINE BRYAN, NICOLA SMETHURST, SARAH BYFORD and BARBARA BARRETT

Background Treatment guidelines identify few adequately powered trials to guide recommendations for anorexia nervosa.

Aims To evaluate the effectiveness of three readily available National Health Service treatments for adolescents (aged 12–18 years) with anorexia nervosa.


Results Each group made considerable progress at 1 year, with further improvement by 2 years. Full recovery rates were poor (33% at 2 years, 27% still with anorexia nervosa). Adherence to in-patient treatment was only 50%. Neither in-patient nor specialist out-patient therapy demonstrated advantages over general CAMHS treatment by intention to treat, although some CAMHS out-patients were subsequently admitted on clinical grounds. In-patient treatment (randomised or after out-patient transfer) predicted poor outcomes.

Conclusions First-line in-patient psychiatric treatment does not provide advantages over out-patient management. Out-patient treatment failures do very poorly on transfer to in-patient facilities.

Declaration of Interest None.

Funding detailed in Acknowledgements.

†See pp. 436–440, this issue.

Recent systematic reviews (Gowers & Bryant-Waugh, 2004; National Collaborating Centre for Mental Health, 2004; Treasure & Schmidt, 2004) have drawn attention to the shortage of high-quality, adequately-powered treatment trials for anorexia nervosa. Indeed, the National Institute for Clinical Excellence (NICE) evidence-based guideline (National Collaborating Centre for Mental Health, 2004) was unable to make a single Grade A treatment recommendation across the age range. The choice of treatment setting has tended to be based on clinical judgement and the availability of different models of service rather than research evidence (National Collaborating Centre for Mental Health, 2004). Debate about the merits of in-patient management frequently fails to distinguish between (often brief) medical admission and longer psychiatric admission aimed at a combination of weight restoration, normal eating and psychological change. We report here a large population-based randomised controlled trial (RCT) of the three main treatments available for adolescents in the UK in order to clarify the relative merits of in-patient psychiatric treatment and two forms of out-patient management.

METHOD

The Treatment Outcome for Child and adolescent Anorexia Nervosa (TOuCAN) trial aimed to compare the clinical effectiveness of in-patient against specialist out-patient and treatment as usual in the community. The study also examined the cost-effectiveness of each approach (Byford et al, 2007, this issue) and user satisfaction with each treatment (not reported here).

Our hypotheses were that: (a) the more intensive in-patient treatment would be more effective than out-patient treatment; and (b) specialist out-patient treatment would be more effective than general child and adolescent mental health service (CAMHS) treatment.

Participants

The trial took place in the north-west of England. The population (total 7.2 million) is served by 38 community CAMHS and four in-patient psychiatric units. The study aimed to recruit as complete a series as possible of consecutive cases referred to community CAMHS. In total 35 out of 38 CAMHS services agreed to refer to the trial.

Inclusion criteria were male or female adolescents aged 12–18 years with a diagnosis of anorexia nervosa according to DSM-IV criteria (American Psychiatric Association, 1994) modified for this age group as follows: food restriction with or without compensatory behaviours; weight below 85% of that expected within 1 month of assessment, based on age and current height or previous height centile; intense fear of gaining weight or undue influence of weight or shape on self-evaluation; primary or secondary amenorrhoea of at least 3 months, or menstruation only while on the contraceptive pill. No exclusions were made on grounds of clinical severity, but the responsible clinician reserved the right to refer for acute medical management if required. Those with severe intellectual disability and severe, chronic comorbid physical conditions affecting digestion or metabolism were excluded.

Recruitment strategy

Child and adolescent mental health services identified patients with probable anorexia nervosa and invited them to meet the researchers. The research team supported by a clinician then interviewed the young person (generally with a parental informant), confirmed the diagnosis and obtained informed consent for them to take part in the randomisation, along with baseline measures. Those agreeing were sent an appointment at the allocated treatment facility closest to their home. The recruitment and consent strategy was approved by the North-West Multi-Centre Research Ethics Committee. Treatment allocation was carried out by an independent randomisation service using stochastic minimisation controlling for gender, age above and below 16 years and body mass index (BMI) above and below 15.5.
Treatments

In-patient psychiatric treatment (4 services)

This was provided within generic children’s or adolescent psychiatric in-patient units. All four services had substantial experience in treating eating disorders, although they were not exclusively eating disorder services. In keeping with the national census findings (O’Herlihy et al., 2003), anorexia nervosa often comprised the most prevalent diagnosis within the units. Treatment lasted 6 weeks in the first instance, extended as clinically indicated and determined by the treating service. The treatment was not manualised, but services met at the outset to identify core elements in treatment. They all used a multidisciplinary psychiatric approach with the aim of normalising eating, restoring healthy weight and facilitating psychological (cognitive) change. Each participant received both individual supportive or cognitive therapies and family therapy. There was a high expectation of early behavioural change and services employed a weight restoration programme with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attended the unit school subject with an expected weight increase of 800–1000 g per week. Participants were ambulant and attende...
check the robustness of the analysis (Efron & Tibshirani 1993). Diagnostic outcome category was modelled using ordinal logistic regression with the same covariates (McCullagh, 1980).

**RESULTS**

**Participants**

The 35 CAMHS identified 347 young people between 2000 and 2003; 100 were excluded because they did not fulfil the diagnosis (n=98, chiefly eating disorder not otherwise specified; EDNOS), and because of physical comorbidity (cystic fibrosis n=1, epilepsy n=1); 31 young people refused all consent, and 46 agreed to follow-up but not randomisation; 170 were randomised to treatment, but 3 were excluded post-randomisation because information subsequently came to light that questioned the diagnosis (2 reclassified EDNOS, 1 chronic fatigue syndrome). The final sample comprised 216 young people, of whom 167 were randomised and are reported here (see Fig. 1 for CONSORT details).

Annual CAMHS audits were carried out to identify young people known to them but not referred to the trial. This identified a further 25 young people, chiefly where the diagnosis had emerged after some time of CAMHS treatment. Overall the study recruited 79% of young people with anorexia nervosa known to community CAMHS in this period (215 out of 271).

**Demographic characteristics**

Participants were aged between 11 years 11 months and 17 years 11 months (mean 14 years 11 months); 153 (92%) were female; 127 (76%) experienced the restricting subtype and 40 (24%) the binge purging subtype of anorexia nervosa. Mean length of history was 13 months; 104 (62.3%) were EDNOS, 1 chronic fatigue syndrome). The remainder had no treatment, 4 opted for other treatment in the initial phase, and no other treatment in the initial phase, whereas the Morgan–Russell scales indicate greater difficulty, whereas the Morgan–Russell scales indicate greater clinical severity by a lower score.

**Clinical features**

Tables 1 and 2 show the presenting features according to allocated treatment. The treatment groups were generally moderately to severely ill (mean weight for height 78.0%, lowest 57.9%). Eight had a weight for height above the anorexic threshold. Of these, 4 were included because they lost significant weight in the 4 weeks following assessment, or they had previously attained a greater height percentile, suggesting stunting of growth, whereas 4 others with borderline weights were included because they fulfilled the other criteria plus significant (>15% and generally >20%) weight loss with amenorrhoea. Five females were sporadically menstruating, but at lower than 85% weight for height. There were no significant differences between groups on any variable; including length of history. For the EDI, MFQ, FAD and HoNOSCA a higher score indicates greater difficulty, whereas the Morgan–Russell scales indicate greater clinical severity by a lower score.

**Adherence to treatment allocation**

Adherence to allocated treatment was 65% but varied between groups. For in-patient treatment, defined as a 4-week in-patient stay, 28 out of 57 adhered (49.1%). In most cases, those failing to adhere agreed initially to admission and then bargained their way out by achieving a small weight gain in the short time between randomisation and admission. Mean length of stay for those admitted was 15.2 weeks. For specialist out-patient treatment, defined as a minimum of 6 attendances, 41 out of 55 adhered (74.5%). Of the remainder, 10 changed their mind and opted for general CAMHS treatment (generally because of travelling distance), 3 were admitted before treatment could start and 1 dropped out of all treatment. For general CAMHS treatment, defined as attending general CAMHS and no other treatment in the initial phase, 38 out of 55 adhered (69.1%). Two of the remainder had no treatment, 4 opted for specialist out-patient treatment and 11 referred to an alternative by clinician preference (10 in-patient, 1 specialist outpatient).

**Clinical outcomes**

Every participant was traced, with the main outcome measures completed as follows: diagnostic outcome and outcome category, 164 participants (98%) at 1 year, 160...
## Table 1
Intention-to-treat analysis of 1- and 2-year outcomes for Morgan–Russell Average Outcome Scale

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Specialised out-patient</th>
<th>General out-patient</th>
<th>In-patient</th>
<th>In-patient – out-patient</th>
<th>General – specialist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>n</td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td><strong>Sub-scale A (food intake)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>3.4</td>
<td>2.0</td>
<td>55</td>
<td>3.1</td>
<td>1.6</td>
</tr>
<tr>
<td>1 year</td>
<td>7.4</td>
<td>2.8</td>
<td>52</td>
<td>7.8</td>
<td>2.6</td>
</tr>
<tr>
<td>2 years</td>
<td>8.1</td>
<td>2.6</td>
<td>51</td>
<td>7.9</td>
<td>2.8</td>
</tr>
<tr>
<td><strong>Sub-scale B (menstruation)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.8</td>
<td>2.1</td>
<td>50</td>
<td>1.0</td>
<td>2.4</td>
</tr>
<tr>
<td>1 year</td>
<td>5.4</td>
<td>5.1</td>
<td>45</td>
<td>5.6</td>
<td>5.3</td>
</tr>
<tr>
<td>2 years</td>
<td>7.1</td>
<td>5.4</td>
<td>42</td>
<td>7.2</td>
<td>5.5</td>
</tr>
<tr>
<td><strong>Sub-scale C (mental state)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>5.2</td>
<td>1.9</td>
<td>55</td>
<td>5.3</td>
<td>2.2</td>
</tr>
<tr>
<td>1 year</td>
<td>7.3</td>
<td>2.9</td>
<td>52</td>
<td>7.3</td>
<td>2.9</td>
</tr>
<tr>
<td>2 years</td>
<td>8.1</td>
<td>2.8</td>
<td>51</td>
<td>8.0</td>
<td>3.2</td>
</tr>
<tr>
<td><strong>Sub-scale D (psychosexual adjustment)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>5.8</td>
<td>2.8</td>
<td>55</td>
<td>5.5</td>
<td>2.7</td>
</tr>
<tr>
<td>1 year</td>
<td>7.9</td>
<td>3.6</td>
<td>51</td>
<td>7.8</td>
<td>3.3</td>
</tr>
<tr>
<td>2 years</td>
<td>8.8</td>
<td>3.2</td>
<td>51</td>
<td>8.2</td>
<td>3.5</td>
</tr>
<tr>
<td><strong>Sub-scale E (socio-economic status)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>7.2</td>
<td>3.0</td>
<td>55</td>
<td>7.8</td>
<td>2.0</td>
</tr>
<tr>
<td>1 year</td>
<td>8.6</td>
<td>3.2</td>
<td>52</td>
<td>9.3</td>
<td>2.7</td>
</tr>
<tr>
<td>2 years</td>
<td>9.3</td>
<td>2.7</td>
<td>51</td>
<td>9.6</td>
<td>2.7</td>
</tr>
<tr>
<td><strong>Average Outcome Scale¹</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>4.6</td>
<td>1.5</td>
<td>55</td>
<td>4.7</td>
<td>1.3</td>
</tr>
<tr>
<td>1 year</td>
<td>7.3</td>
<td>2.3</td>
<td>52</td>
<td>7.6</td>
<td>2.2</td>
</tr>
<tr>
<td>2 years</td>
<td>8.4</td>
<td>2.4</td>
<td>51</td>
<td>8.3</td>
<td>2.6</td>
</tr>
</tbody>
</table>

1. Statistical analysis adjusted for baseline age, gender, site and baseline Mood and Feelings Questionnaire score.
2. Mean of five sub-scales.
Table 2  Intention-to-treat analysis for secondary outcomes at 1 and 2 years

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Specialised out-patient</th>
<th>General out-patient</th>
<th>In-patient</th>
<th>In-patient – out-patient</th>
<th>General – specialist</th>
</tr>
</thead>
</table>
|                    | Mean s.d. n             | Mean s.d. n         | Mean s.d. n | Mean s.d. n              | Difference1 (95% CI) | P
| Body mass index    |                         |                     |            |                          |                     |
| Baseline           | 15.3 1.6 55             | 15.5 1.6 55         | 15.3 1.6 57 | -0.41 (-1.18 to 0.35)    | 0.28 0.35 (-0.55 to 1.25) | 0.45 |
| 1 year             | 17.9 2.2 52             | 18.3 2.7 50         | 17.5 2.2 52 | -0.38 (-1.22 to 0.47)    | 0.38 0.48 (-0.53 to 1.50) | 0.35 |
| 2 years            | 18.7 2.1 50             | 19.4 2.7 48         | 18.7 2.8 52 |                         |                     |
| Weight for height  |                         |                     |            |                          |                     |
| Baseline           | 77.1 8.1 55             | 78.8 7.9 55         | 78.2 8.1 57 |                         |                     |
| 1 year             | 88.2 10.6 52            | 90.9 13.3 50        | 86.7 9.9 52 | -2.29 (-6.01 to 1.44)    | 0.23 1.39 (-2.93 to 5.74) | 0.53 |
| 2 years            | 90.1 9.8 50             | 94.2 13.0 46        | 90.4 13.3 51 | -1.63 (-5.74 to 2.47)    | 0.43 2.72 (-2.17 to 7.62) | 0.27 |
| EDI Total          |                         |                     |            |                          |                     |
| Baseline           | 86.5 47.5 54            | 88.5 51.4 52        | 89.6 44.5 56 |                         |                     |
| 1 year             | 57.6 54.0 44            | 69.4 53.3 45        | 60.6 52.9 43 | -5.80 (-23.36 to 11.77)  | 0.52 6.26 (-14.35 to 26.88) | 0.55 |
| 2 years            | 52.5 49.1 42            | 61.0 52.0 40        | 40.3 36.4 43 | -14.10 (-30.75 to 2.55)  | 0.10 8.41 (-11.32 to 28.15) | 0.40 |
| FAD                |                         |                     |            |                          |                     |
| Baseline           | 2.12 0.53 54            | 2.13 0.59 52        | 2.08 0.5 56 | -0.07 (-0.25 to 0.10)    | 0.41 -0.12 (-0.33 to 0.08) | 0.23 |
| 1 year             | 2.08 0.55 45            | 1.97 0.57 46        | 1.95 0.5 43 |                         |                     |
| 2 years            | 1.99 0.59 39            | 2.02 0.65 41        | 1.99 0.5 42 | 0.04 (-0.17 to 0.24)     | 0.72 0.05 (-0.19 to 0.29) | 0.68 |
| MFQ                |                         |                     |            |                          |                     |
| Baseline           | 30.1 14.7 54            | 32.4 16.1 53        | 32.6 14.6 56 |                         |                     |
| 1 year             | 19.3 16.7 46            | 23.8 17.7 46        | 18.2 15.6 43 | -4.35 (-10.07 to 1.38)   | 0.14 2.73 (-3.82 to 9.27) | 0.41 |
| 2 years            | 17.1 15.1 42            | 24.2 20.2 42        | 15.8 14.5 42 | -3.58 (-9.54 to 2.37)    | 0.24 5.15 (-1.65 to 11.94) | 0.14 |
| HoNOSCA, clinician-rated |                   |                     |            |                          |                     |
| Baseline           | 20.7 7.5 55             | 20.0 5.7 55         | 20.0 5.6 57 | -1.20 (-4.07 to 1.68)    | 0.41 -1.34 (-4.70 to 2.01) | 0.43 |
| 1 year             | 16.8 9.7 49             | 15.0 9.1 53         | 14.2 7.4 52 |                         |                     |
| 2 years            | 13.7 8.9 51             | 13.8 9.8 52         | 14.3 9.1 52 | 1.06 (-2.00 to 4.12)     | 0.49 0.19 (-3.38 to 3.76) | 0.92 |
| HoNOSCA, self-rated |                     |                     |            |                          |                     |
| Baseline           | 17.4 9.9 53             | 16.5 10.0 54        | 15.6 9.5 53 |                         |                     |
| 1 year             | 11.7 9.0 44             | 10.5 10.0 45        | 8.6 8.2 42 | -1.65 (-4.99 to 1.68)    | 0.33 -2.13 (-5.87 to 1.64) | 0.26 |
| 2 years            | 8.9 8.1 43              | 10.0 9.8 37         | 7.7 8.6 43 | -0.65 (-3.83 to 2.54)    | 0.69 1.15 (-2.53 to 4.83) | 0.54 |

EDI, Eating Disorder Inventory 2; FAD, Family Assessment Device; MFQ, Mood and Feelings Questionnaire; HoNOSCA, Health of the Nation Outcome Scale for Children and Adolescents.

1. Statistical analysis adjusted for baseline age, gender, site and baseline MFQ score.
Table 3  Categorical outcomes at 1 and 2 years by intention to treat

<table>
<thead>
<tr>
<th></th>
<th>1-year outcome, n(^1)</th>
<th>2-year outcome, n(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Good</td>
<td>Intermediate</td>
</tr>
<tr>
<td>General CAMHS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherers (n=38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Subsequently admitted</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Non-adherers (n=17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated as out-patient</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Treated as in-patient</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Untreated</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>10 (18)</td>
<td>31 (56)</td>
</tr>
<tr>
<td>Specialist out-patient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherers (n=41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Subsequently admitted</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Non-adherers (n=14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated as out-patient</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Treated as in-patient</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>8 (15)</td>
<td>22 (40)</td>
</tr>
<tr>
<td>In-patient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherers (n=28)</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Non-adherers (n=29)</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>12 (21)</td>
<td>18 (32)</td>
</tr>
</tbody>
</table>

1. Ordinal logistic regression, 1 year \(P=0.22\), 2 years \(P=0.89\).

(96%) at 2 years; MRAOS, 157 (94%) at 1 year, 155 (93%) at 2 years; BMI/weight for height, 154 (92%) at 1 year, 150 (90%) at 2 years; HoNOSCA 154 (92%) at 1 year, 153 (93%) at 2 years. These were achieved by face-to-face interview in 129 participants (77%) at 1 year and 121 (73%) at 2 years. Outcome data were obtained by telephone interview or interview with a health professional informant in 34 (20%) at 1 year and 40 (24%) at 2 years. The remaining 4 at 1 year and 6 at 2 years were all traced (alive) but little or no information was obtained on their health status.

Table 3 shows the categorical outcomes based on those employed in the Maudsley studies (Russell et al, 1987) and we employed a high threshold for assigning recovery. A good outcome indicates a full recovery from anorexia nervosa (weight above 85% of expected, return of menstruation, binging/purging no greater than once per month). A poor outcome was indicated if weight was not above 85% or the young person was still being treated as an in-patient for anorexia nervosa. The intermediate category comprises those whose weight had risen to within the normal range, but without return of menstruation, with binging/purging at a frequency greater than monthly or considerable residual concerns about weight and shape according to Morgan–Russell scale A scores for food intake.

Reliability of assessment measures

Interrater reliability (IRR) series were carried out within research site and between sites at baseline; intraclass correlation (ICC) coefficients were as follows: MRAOS, Manchester 0.93, Liverpool 0.97, intersite 0.96, IRR at 1 year 0.93, 2 years 0.90; HoNOSCA, Manchester 0.83, Liverpool 0.98, intersite 0.87, IRR at 1 year 0.89, 2 years 0.89.

1 year outcome

All groups made substantial mean improvements in terms of weight, global measures and self-reported psychopathology (Tables 1 and 2). In an intention-to-treat analysis there are no statistically significant differences between the three groups. In particular, the mean values on the MRAOS are remarkably similar across the treatments. Those allocated to general CAMHS treatment were less likely to still have anorexia nervosa at 1 year (Table 3), but by intention to treat there were no significant differences between the three groups (ordinal logistic regression \(P=0.22\)). For the two out-patient treatment arms there was a much better outcome for those who fully adhered to treatment compared with those failing to adhere or later transferring away from allocated treatment. Specifically, for general CAMHS treatment, only 1 out of 17 admitted for in-patient treatment had a good outcome at 1 year, whereas for the specialist out-patient programme, none of the 14 who initially failed to adhere to the allocated programme had a good outcome, nor any of 14 subsequently admitted to in-patient treatment.

Sub-analysis of those allocated to in-patient treatment

The relatively poor outcomes at 1 year of those allocated to in-patient treatment merits further exploration. Adherence to treatment was poor (49%, 28 out of 57) for this option, therefore in theory this might have compromised the effectiveness
DISCUSSION

This trial reports the outcome of a large, population-based RCT of adolescents with anorexia nervosa. Although frequently a chronic condition, treatment in a range of services brought about significant improvement by 1 year after presentation, with further progress by 2 years. Fewer than 1 in 5 fully recovered within 1 year, but one-third had recovered by 2 years, with only a quarter still having anorexia nervosa at this time point.

Contrary to our hypotheses, there was no advantage for specialist over general CAMHS treatment or in-patient over outpatient management. It could be argued that the in-patient services in the study were not truly specialised, as they were not exclusive eating disorder facilities. However, all four units had extensive experience and tradition of treating such patients. Indeed 17 adolescents entered other (often exclusive) specialist in-patient services in the follow-up period of the study and still had poor outcomes at 2 years. Most of those who found their way into in-patient management had failed to improve with outpatient treatment. Nevertheless, their poor outcomes challenge the intuitive clinical belief that a step up progression from outpatient to in-patient psychiatric care is indicated for those who fail to make progress. The outcomes of those allocated to and receiving in-patient management were also rather poor. It may be that the decision to accept randomised admission is based on a number of negative prognostic variables, rather than it reflecting on the in-patient treatment itself. Our analysis suggests that those agreeing to admission were marginally thinner and had a lower MRAOS sub-scale A (food intake) score. However, given that the presenting values on these variables did not account for the differences in outcome, some unmeasured variables such as motivation or family resources may have accounted for the differences in response. This finding does not deny the necessity of emergency medical management of physical complications in an in-patient setting, which may on occasions be life-saving, but our results do suggest that in-patient management rarely leads to comprehensive recovery, as opposed to improvement or stability within the condition. The health economic implications of this finding are presented elsewhere, (Byford et al., 2007, this issue), but this finding has significant cost implications.

Comparison with previous research

The NICE eating disorder guideline (National Collaborating Centre for Mental Health, 2004) highlighted the shortage of quality treatment trials for anorexia nervosa. A number of relatively small RCTs (Russell et al., 1987; Le Grange et al., 1992; Eisler et al., 1997; Robin et al., 1999; Eisler et al., 2000; Lock et al., 2003) have suggested promising outcomes of family interventions for adolescents and few would contest the necessity of involving parents in their treatment. Most of this research has followed the Maudsley model, but differences in research design (for example inclusion in some studies of participants who have had their weight restored) makes for uncertainties in interpretation, particularly as this treatment has not been fully tested against other approaches. The present study was devised before the more recent positive outcomes of family-based treatment were published and it is of note that our findings suggest poorer outcomes. We had been impressed by the preliminary outcomes of extended CBT (Fairburn et al., 2003) in addressing the core psychopathology of eating and weight concerns and questioned the power of family-based treatment to address these as opposed to behavioural aspects of the condition. Recent research from the Maudsley group (Eisler et al., 2000) has cast some doubt on the value of conjoint family therapy in a trial that found it less effective than separated family therapy, based on the Morgan–Russell outcome categories. Strikingly no young people had a good outcome where there was high expressed emotion in the family. Clearly further adequately powered studies are required of family-based treatment against CBT, either as described here or the transdiagnostic form devised by the Oxford group (Fairburn et al., 2003).

Treatment setting has been investigated far less. The relatively underpowered St George’s trial failed to find an advantage for in-patient over specialist out-patient treatment in a mixed age sample (Crisp et al., 1991, Gowers, 1991). This led Madsen et al. (2001) to conclude in their systematic review that out-patient treatment in a specialist eating disorder service was as effective as in-patient treatment in those not so severely ill as to warrant emergency admission. Furthermore, these reviewers estimated the costs of out-patient treatment to be approximately one-tenth the cost of in-patient treatment.
The facility to offer long-term psychiatric (as opposed to medical) treatment in the UK has often been highly valued, clinical intuition suggesting that more intensive treatments should be more effective than briefer, non-specialist treatment for a condition that is often chronic and has a high morbidity and mortality (Herzog, 1992). However, there is little research evidence to demonstrate the benefits of lengthy in-patient psychiatric treatment and a recent survey (Roots et al, 2006) of adolescent services in the UK and Europe revealed great variation in typical length of stay, target weight and treatment philosophy. The St George’s (Crisp et al, 1991) and Maudsley studies (Russell et al, 1987) showed that although the majority of those receiving lengthy in-patient treatment gained weight to normal levels, many had lost a significant amount of weight by 1 year follow-up. Furthermore, there has been little or no research into potential unwanted effects of different treatment settings. The poor outcomes of in-patient treatment in a naturalistic cohort of young people have led to speculation that certain features of anorexia nervosa, for example ineffectiveness, low self-esteem, interpersonal distrust, might be exacerbated by lengthy admission (Gowers et al, 2000).

Research into the length and content of out-patient treatment has also provided mixed findings. Clinical intuition would suggest that intensive, longer-term treatment is required for this condition. However, one study (McIntosh et al, 2005) recently reported that supportive clinical management was more effective (in adults) than two specialised and intensive forms of psychotherapy, although these authors now describe this as a specialist treatment (McIntosh et al, 2006). A further trial (Lock et al, 2005) found that a short course of family therapy appeared to be as effective as a long course for adolescents with short-duration anorexia nervosa. There was a suggestion, however, that those with more severe obsessive-compulsive thinking and non-intact families benefited from longer treatment. It is clear that emerging research findings are challenging established beliefs about the treatment of this condition and further clarification is required. There has been much recent interest in the importance of patient motivation and interventions aimed at improving it (Vitousek et al, 1998; Geller et al, 2001), in order to overcome resistance or passive acceptance of treatment.

SIMON G. GOWERS, FRCPsych, MPhil, Section of Adolescent Psychiatry, University of Liverpool; ANDREW CLARK, FRCPsych, Bolton Salford and Trafford Mental Health NHS Trust; CHRIS ROBERTS, PhD, Biostatistics Group, Division of Epidemiology and Health Sciences, University of Manchester; ALISON GRIFFITHS, BA, Section of Adolescent Psychiatry, University of Liverpool; VANESSA EDWARDS, BA, Peninsula College of Medicine and Dentistry, Exeter; CLAUDINE BRYAN, BSc, NICOLA SMETHURST, BA, Section of Adolescent Psychiatry, University of Liverpool; SARAH BYFORD, MSc, BARBARA BARRETT, MSc, Kings College London, UK.

Correspondence: Professor S.G. Gowers, Academic Unit, Section of Adolescent Psychiatry, University of Liverpool, Chester CH2 1AW, UK. Email: simon.gowers@cpwnt.nhs.uk

(First received 9 February 2007, final revision 22 June 2007, accepted 11 July 2007)

Together, these findings and the lack of evidence from systematic reviews suggest much uncertainty remains about effective treatment for this condition, although there is a growing literature challenging approaches delivered without the patient’s active cooperation.

Strengths and limitations

This study is much larger than those reported in the literature to date and includes around four-fifths of incident cases known to child and adolescent mental health services in the north-west of England over a 3-year period. We achieved a high follow-up rate with demonstrably reliable outcome measures. The outcome of our individual CBT was poorer than reported for family-based treatment – a direct comparison is required on a similar population to clarify this further. Not all participants fully adhered to randomised treatment. It is, on the one hand a problem for long-term follow-up of RCTs of chronic conditions that some will subsequently engage in other treatments on clinical grounds. On the other hand, this provides an insight into use of services and enables evaluation of the outcome of step-up treatments.

ACKNOWLEDGEMENTS

This project is funded by the NHS (E) HTA programme, grant no. 97/42/02, and supported by the North West MMRN. Trial registry: ISRCTN19345394. They are responsible for approving the design and supervising the conduct of the study.

We thank the child and adolescent mental health services in the north-west of England who contributed to the project and especially the Young People’s Centre Chester and the Adolescent Psychiatry Service, Prestwich. The following also made significant contributions: Dr Asha Bhat Clinical Lead, Specialist Out-Patient Programme, Manchester; Sheeba Ehsan, Irene Cohen, Laura Rowlands, Dr Tania Stanway, Kandy Giles, Rachel Charlesworth, Matilda Moffatt and Sarah Bailey-Rogers.

REFERENCES


Economic evaluation of a randomised controlled trial for anorexia nervosa in adolescents†

SARAH BYFORD, BARBARA BARRETT, CHRIS ROBERTS, ANDREW CLARK, VANESSA EDWARDS, NICOLA SMETHURST and SIMON G. GOWERS

Background Young people with anorexia nervosa are often admitted to hospital for treatment. As well as being disruptive to school, family and social life, in-patient treatment is expensive, yet cost-effectiveness evidence is lacking.

Aims Cost-effectiveness analysis of three treatment strategies for adolescents with anorexia nervosa.

Method UK multicentre randomised, controlled trial comparing in-patient psychiatric treatment, specialist out-patient treatment and general out-patient treatment. Outcomes and costs assessed at baseline, 1 and 2 years.

Results There were 167 young people in the trial. There were no statistically significant differences in clinical outcome between the three groups at 2 years. The specialist out-patient group was less costly over the 2-year follow-up (mean total cost £26 738) than the in-patient (£34 531) and general out-patient treatment (£40 794) groups, but this result was not statistically significant. Exploration of the uncertainty associated with the costs and effects of the three treatments suggests that specialist out-patient treatment has the highest probability of being cost-effective.

Conclusions On the basis of cost-effectiveness, these results support the provision of specialist out-patient services for adolescents with anorexia nervosa.

Declaration of interest None. Funding detailed in Acknowledgements.

†See pp. 427–435, this issue.

Anorexia nervosa is commonly associated with severe physical, psychological and social impairments, high levels of mortality (National Collaborating Centre for Mental Health, 2004) and a significant cost burden (Striegel-Moore et al, 2000; Simon et al, 2005). Young people with anorexia nervosa are often admitted to hospital for treatment. This is disruptive to school, family and social life, and in-patient treatment is an expensive option, yet evidence to support its cost-effectiveness is lacking (Romeo et al, 2005). A recent systematic review did not identify any economic evaluations of treatments for anorexia nervosa (National Collaborating Centre for Mental Health, 2004). One subsequent modelling study has been reported (Crow & Nyman, 2004). However, the authors acknowledge the limitations of modelling and the need for controlled trials. We report the results of an economic evaluation of psychiatric in-patient, specialist out-patient and general out-patient services for adolescents with anorexia nervosa carried out alongside a randomised controlled clinical evaluation – the Treatment Outcome for Child and Adolescent Anorexia Nervosa (TOuCAN) trial (ISRCTN39345394).

METHOD

The aim of the TOuCAN trial was to explore the clinical and cost-effectiveness of in-patient, specialist out-patient and general out-patient services for adolescents with anorexia nervosa. The main economic hypotheses were that: (a) specialist out-patient services would be more cost-effective than general out-patient treatment in community child and adolescent mental health services (CAMHS); and (b) out-patient services would be more cost-effective than in-patient services.

Clinical outcomes

Research assessors masked to treatment allocation carried out assessments at baseline, 1 and 2 years after trial entry. The a priori primary outcome measure for the clinical and economic evaluation was the Morgan–Russell Average Outcome Scale (MRAOS; Morgan & Hayward, 1988), adjusted for adolescents. Full details of outcome measures, their reliability and validity, are reported in the accompanying paper (Gowers et al, 2007, this issue).

Cost

The economic evaluation took a broad service-providing perspective, including that of the health, social services, education,
voluntary and private sectors. Information on resource use was collected in interview at the 1- and 2-year follow-up assessments using the Child and Adolescent Service Use Schedule (CA–SUS), developed by the authors in previous research with young people and adapted for the purpose of the current study (Byford et al., 1999; Harrington et al., 2000; Barrett et al., 2006). Data on hospital contacts were collected from clinical records to avoid patients revealing their treatment group to the research assessors.

All unit costs were for the financial year 2003–2004. Costs in the second year were discounted at a rate of 3.5%, as recommended by the National Institute for Health and Clinical Excellence (National Institute for Clinical Excellence, 2004). This was varied from 0 to 6% in sensitivity analysis.

All National Health Service (NHS) hospital contacts, including the trial interventions, were costed using NHS reference costs (Department of Health, 2004). The unit costs of private sector in-patient stays were collected through direct personal communication with each facility. Unit costs of community health and social services were taken from national publications (Curtis & Netten, 2004). The costs of schooling came from a number of sources including various Ofsted reports (the inspectorate and regulatory body for schools in England, see http://www.ofsted.gov.uk) and published documents (Berridge et al., 2002; Independent Schools Council, 2004). Medications were costed using the British National Formulary (British Medical Association & Royal Pharmaceutical Society of Great Britain, 2004). Where necessary, unit costs were inflated to 2003–2004 costs using the Hospital and Community Health Services inflation indices (Curtis & Netten, 2004).

Statistical methods

All economic analyses were carried out on an intention-to-treat basis using a statistical analysis plan drawn up prior to the analysis of the data. The primary analysis was of total costs over 2 years for the sample of young people with complete economic data.

Although costs were not normally distributed, analyses compared mean costs in the three groups using analysis of co-variance with covariates for pre-specified baseline characteristics: site (Liverpool and Manchester), gender, age at baseline, baseline BMI and baseline MRAOS score. The robustness of the parametric tests was confirmed using bootstrapping (Efron & Tibshirani, 1993), as recommended by Barber & Thompson (1998). The impact of drop-out was assessed by comparing the baseline characteristics of participants who had missing data with those who had full economic data.

Cost-effectiveness was assessed through the calculation of incremental cost-effectiveness ratios (ICER) – the additional costs of one intervention compared with another divided by the additional effects of one intervention compared with another (Van Hout et al., 1994), in this case using the MRAOS measure of effectiveness. When more than two strategies are compared, ICERs are calculated using rules of dominance and extended dominance (Johannesson & Weinstein, 1993). Strategies are ranked by cost, from the least expensive to the most expensive, and if a strategy is more expensive and less effective than the previous strategy, it is said to be dominated and is excluded from the calculation of ICERs. This process compares strategies in terms of observed differences in costs and effects, regardless of the statistical significance of the difference.

Uncertainty around the cost and effectiveness estimates was represented by cost-effectiveness acceptability curves (Van Hout et al., 1994; Fenwick et al., 2001). Repeat re-sampling from the costs and effectiveness data (bootstrapping) was used to generate a distribution of mean costs and effects for the three treatments. These distributions were used to calculate the probability that each of the treatments is the optimal choice, subject to a range of possible maximum values (ceiling ratio, λ) that a decision-maker might be willing to pay for a unit improvement in MRAOS score. Cost-effectiveness acceptability curves are presented by plotting these probabilities for a range of possible values of the ceiling ratio. These curves incorporate the uncertainty that exists around the estimates of mean costs and effects as a result of sampling variation and uncertainty regarding the maximum cost-effectiveness ratio that a decision-maker would consider acceptable (Fenwick & Byford, 2005).

Missing data were explored in three sensitivity analyses using the following data: (a) hospital cost data collected from clinical records and available for a larger sample of young people than full economic data from the CA–SUS; (b) hospital cost data collected from records plus missing non-hospital cost data imputed using the last value carried forward approach for participants with missing year-2 data; and (c) hospital cost data collected from records plus mean imputation by randomised group of missing non-hospital cost data. The results of all sensitivity analyses are reported in the data supplement to the online version of this paper.

RESULTS

Participants

There were 167 young people entering the trial and they were randomised to in-patient care (n=57), specialist out-patient services (n=55) or treatment as usual by general CAMHS (n=55). Full economic data for the 2-year follow-up period were available for 135 young people (81%), 47 in the in-patient group, 45 in the specialist out-patient group and 43 in the general out-patient group. A comparison of baseline characteristics (site, age, gender, BMI and MRAOS score) revealed no significant differences between those included in the economic evaluation and those who were missing, and there was no difference overall in missing data between the three treatment groups. Length of follow-up varied somewhat (range 99–118 weeks), however there was no significant difference in length of follow-up between the three treatment groups (mean 105 weeks in the in-patient and general out-patient groups and 106 in the specialist out-patient group).

Outcomes

There were no significant differences between the three groups by intention to treat at either 1- or 2-year follow-up on the MRAOS (in-patient 2-year global score 8.3, specialist out-patient 8.4, general out-patient 8.3; P=0.838). Full clinical outcome data are reported by Gowers et al (2007, this issue).

Resource use

Table 1 details the mean number of contacts young people had with all services over the 2-year follow-up period. Resource use differed little between the groups except for in-patient and out-patient contacts. The general out-patient group spent more time in hospital and had a greater number of out-patient attendances on average than the specialist out-patient or in-patient groups. The specialist out-patient group spent the least amount of time in hospital.
Exploration of hospital contacts over time reveals that a larger proportion of days were spent in hospital in the first year (in-patient group 62 days, specialist out-patient 35, general out-patient 65) than the second year (in-patient group 12 days, specialist out-patient 20, general out-patient 24). The in-patient treatment of those allocated to out-patient treatment generally occurred after assigned treatment had ended. Details of adherence to treatment are given in the accompanying paper (Gowers et al., 2007, this issue).

Hospital contacts reported in Table 1 include all specialties. However, the vast majority of contacts were psychiatric (71% of in-patient admissions and 90% of in-patient days) or paediatric (20% of admissions and 10% of in-patient days). Other specialties (9% of admissions and 0.2% of in-patient days) included gastroenterology, general medicine, haematology, intensive care unit, obstetrics, orthopaedics, plastic surgery and urology.

Time in education was similar across the three groups, however on average participants spent a significant proportion of the 2-year follow-up period out of education (approximately 10 out of the 24 months of follow-up).

**Costs**

Table 2 details the total mean costs per participant over the 2-year follow-up period. There were no statistically significant differences between the three groups. In terms of observed differences, the specialist out-patient group was consistently cheaper than the other two groups and the general out-patient group was the most expensive of the three. The bootstrapped results differed little and are thus not reported here. Hospital costs constitute the greatest proportion of total costs (93% in each group), with few community health and social services being used.

**Cost-effectiveness analysis**

Using the rules of dominance described in the Method section, specialist out-patient treatment (bootstrapped mean cost per participant £26797; bootstrapped mean effect 8.35) dominates the in-patient group (£34371; 8.26) and the general out-patient group (£40520; 8.26) since it is both cheaper and more effective. Figure 1 illustrates the uncertainty associated with the costs and effects of the three treatments at 2 years and demonstrates that if decision-makers were willing to pay nothing for a unit increase in MRAOS score, there is a 78% chance of specialist out-patient services being the most cost-effective strategy, 16% for in-patient services and only 6% for general CAMHS. The probability of specialist out-patient services being the most cost-effective strategy decreases with increasing levels of willingness to pay for gains in effectiveness, levelling out at around 47%, but remains higher than the other two strategies over the full range of willingness to pay values shown, and beyond. Figure 1 suggests that the probability of our first hypothesis being true is high, i.e. that specialist out-patient services are more cost-effective than general out-patient services. Figure 2 depicts the cost-effectiveness acceptability curve for the second hypothesis, i.e. in-patient v. out-patient services (specialist combined with general), and shows that there is a greater probability of out-patient services being more cost-effective than in-patient services for the full range of values of willingness to pay.

**DISCUSSION**

This paper reports the results of the first economic evaluation of alternative strategies for the treatment of anorexia nervosa using primary data collected from a randomised controlled trial.
Table 2  Total cost (£) per young person over the 2-year follow-up period

<table>
<thead>
<tr>
<th>Sector</th>
<th>In-patient (n=47)</th>
<th>Specialist out-patient (n=45)</th>
<th>General out-patient (n=43)</th>
<th>ANOVA 1</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (s.d.)</td>
<td>Minimum</td>
<td>Maximum</td>
<td>Mean (s.d.)</td>
<td>Minimum</td>
</tr>
<tr>
<td>Secondary health care</td>
<td>32 015 (5 541)</td>
<td>69</td>
<td>280 580</td>
<td>24 724 (46 231)</td>
<td>103</td>
</tr>
<tr>
<td>Primary health care</td>
<td>380 (640)</td>
<td>4</td>
<td>2773</td>
<td>385 (873)</td>
<td>0</td>
</tr>
<tr>
<td>Education</td>
<td>2098 (2115)</td>
<td>0</td>
<td>8783</td>
<td>1595 (1456)</td>
<td>0</td>
</tr>
<tr>
<td>Other community</td>
<td>37 (110)</td>
<td>0</td>
<td>513</td>
<td>35 (104)</td>
<td>0</td>
</tr>
<tr>
<td>Total 2-year cost</td>
<td>34 531 (52 439)</td>
<td>86</td>
<td>282 508</td>
<td>26 738 (46 809)</td>
<td>462</td>
</tr>
<tr>
<td>Total cost per week</td>
<td>325 (487)</td>
<td>1</td>
<td>2588</td>
<td>253 (442)</td>
<td>4</td>
</tr>
</tbody>
</table>

1. Adjusted for site, gender, age at baseline, baseline body mass index and baseline Morgan–Russell Average Outcome Scale score.
2. Includes community social, voluntary and private sector services.

Clinical outcomes, resource use and cost

There were no statistically significant differences in clinical outcomes between the three groups, with results demonstrating similar improvements in all groups over time. The specialist out-patient group was observed to be the cheapest of the three groups and the general out-patient group the most expensive. However, these differences were not statistically significant. These findings were robust to changes in the discount rate and in analyses of missing data.

Observed differences in total mean cost per participant were almost entirely due to differences in the length of time spent in hospital. Secondary healthcare costs accounted for over 90% of all costs and, of this, almost 90% was due to in-patient stays. The majority of in-patient stays took place in the first year. Although not randomised to psychiatric in-patient services, the general out-patient service group spent almost as much time in hospital as the in-patient group, suggesting that general CAMHS were less successful at maintaining these young people in the community than specialist out-patient services. With the exception of CAMHS, participants used very few community health and social services. Months in education was similar across the groups on average, but highlighted the significant proportion of time participants spent out of education, presumably as a result of their illness.

At almost £17 000 per year on average, the annual service costs of caring for this group of young people were high. Although much higher than the cost of conditions generally treated in the community, for
example conduct disorder with annual service cost estimates varying between £1300 and £3200 (Harrington et al, 2000; Romeo et al, 2006), this figure is similar to the cost of a cohort of young people treated in child and adolescent psychiatric in-patient wards, estimated to be £24,000 per admission (Green et al, 2007). Although slightly higher than the costs reported in this paper, the mean length of stay was longer at 116 days.

Cost-effectiveness
Specialist out-patient services were found to be the dominant treatment option in terms of incremental cost-effectiveness (more effective and less costly). Exploration of the associated uncertainty supported this finding. In terms of our hypotheses, the data suggest that specialist out-patient services have a higher probability of being cost-effective than general out-patient services and that out-patient services (specialist combined with general) have a higher probability of being cost-effective than psychiatric in-patient services.

Limitations
Despite substantial differences in observed cost data, these differences did not reach statistical significance. This may be due to inadequate sample sizes for the economic evaluation. Sample size calculations were based on the primary outcome measure, the MRAOS. Calculations on the basis of cost or cost-effectiveness were not feasible at the design stage because of the lack of any relevant published cost data. Although acknowledging this limitation, the use of a decision-making approach to the economic evaluation provides probabilistic evidence of the cost-effectiveness of the alternative treatment strategies, given the data currently available. Although larger trials may be considered in future research, this must be balanced against the cost of additional research in a disease area where low prevalence rates necessitate multicentre evaluation. Analysis of patients excluded because of missing economic data did not suggest any bias; patients included in the economic evaluation did not differ significantly from those excluded and there was no evidence to suggest any bias in missing data between the three treatment groups.

ACKNOWLEDGEMENTS
This project is funded by the UK National Health Service Health Technology Assessment programme, Grant No. 97/42/02 and supported by the North West M-HEL. This paper is dedicated to Professor Richard Harrington.

REFERENCES
Antidepressant effects of augmentative transcranial magnetic stimulation

Randomised multicentre trial


Background Repetitive transcranial magnetic stimulation (rTMS) has been proposed as a new treatment option for depression. Previous studies were performed with low sample sizes in single centres and reported heterogeneous results.

Aims To investigate the efficacy of rTMS as augmentative treatment in depression.

Method In a randomised, double-blind, sham-controlled multicentre trial 127 patients with moderate to severe depressive episodes were randomly assigned to real or sham stimulation for 3 weeks in addition to simultaneously initiated antidepressant medication.

Results We found no difference in the responder rates of the real and the sham treatment groups (31% in each) or in the decrease of the scores on the depression rating scales.

Conclusions The data do not support previous reports from smaller samples indicating an augmenting or accelerating antidepressant effect of rTMS. Further exploration of the possible efficacy of other stimulation protocols or within selected sub-populations of patients is necessary.

Declaration of interest None.

Major depression is one of the leading causes of disease burden worldwide (Berton & Nestler, 2006). Its impact on society with respect to human suffering and economic charge is enormous, and is even projected to increase in upcoming decades (Lopez & Murray, 1998). Since the discovery of drugs with antidepressant properties in the 1950s, no essentially innovative treatment strategy has been established for routine clinical use. Resistance to the available treatment strategies is encountered in 15–30% of patients. New treatment approaches are therefore needed. Repetitive transcranial magnetic stimulation (rTMS) was introduced as a promising new treatment option for depression and showed beneficial effects in single-centre studies (Burt et al, 2002; Kozel & George, 2002; Loo & Mitchell, 2005). However, it remains difficult to draw general conclusions about the antidepressant efficacy of rTMS because of heterogeneous study designs, variable stimulation parameters and low sample sizes (Martin et al, 2003).

METHOD

Study design and participants

The aim of this multicentre trial was to evaluate whether the application of rTMS in a routine clinical setting as an additional strategy to standard antidepressant medication would enhance the clinical improvement of depression compared with sham treatment with regard to the number of responders and the decrease in depression rating scores. Psychiatric departments in seven university clinics – Munich (Ludwig-Maximilian University), Regensburg, Rostock, Tübingen, Ulm and Würzburg in Germany, and Vienna in Austria – with experience in transcranial magnetic stimulation studies participated in this randomised double-blind placebo-controlled, multicentre trial. Randomisation to the real and sham treatment conditions was performed centrally prior to the study by the Institute of Biometrics of the University of Ulm. Patients, raters and medical staff at the in-patient units were all masked to the treatment conditions. The principal investigator (U.H.) at the University of Zürich was responsible for study coordination and central data collection. The Institute of Biometrics of the University of Ulm performed the statistical analysis. All patients gave written informed consent. The study was conducted according to the latest version of the Declaration of Helsinki and was approved by the local ethics committees in each centre.

Inclusion criteria were age 18–75 years; a moderate or severe major depressive episode meeting ICD–10 and DSM–IV criteria (World Health Organization, 1992; American Psychiatric Association, 1994), including bipolar affective disorder, assessed with the Structured Clinical Interview for DSM–IV Axis I Disorders (SCID; First et al, 1998); and a score of 18 or more points on depression rating scales: the Beck Depression Inventory (BDI; Beck et al, 1961), the 21-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960) and the Montgomery–Asberg Depression Rating Scale (MADRS; Montgomery & Åsberg, 1979). The cut-off at 18 points was chosen because in all three scales it is within the range of the transition from mild to moderate severity of depression. Exclusion criteria were neurological and severe medical disorders, psychiatric disorders other than depression, history of epileptic seizures, brain lesions or neurosurgery, cardiac pacemaker, inability to give informed consent, and involuntary hospitalisation. Included patients were given an identification number linked to a centralised computer-generated randomisation code determining real or sham stimulation condition. Randomisation was stratified for centre and for HRSD score >30 or ≤30 at enrolment. Raters underwent training at the beginning of the study to increase interrater reliability.

The following individual and clinical features at baseline were documented (see Table 1): duration of the current episode before rTMS, number of episodes in the history including the current episode (1–3 vs. >3), treatment resistance (no response to two different antidepressant medications and one combination treatment with treatment periods of at least 4 weeks each in sufficient dosage for the current episode), polarity (depressive episode within unipolar...
or bipolar disorder), a medical record of family history for depression, and history of a severe psychosocial stressor in the year before manifestation of the current episode (such as death of a close relative, separation from a partner or loss of work).

**Transcranial magnetic stimulation**

Each clinic used the locally available magnetic stimulator with figure-of-eight coils: the Magstim Rapid (Magstim Company Ltd, Whitland, UK; double 70 mm coil, P7N 9790) in Munich, Tubingen, Vienna and Regensburg; the Medtronic Magpro (Medtronic Inc., Minneapolis, USA; coil MC-B70) in Ulm and Wurzburg; and the Medtronic Maglite r25 (Medtronic Inc., Minneapolis, USA; coil MC-B70) in Rostock. A biphasic pulse waveform was selected for all stimulations. The participant was seated in a comfortable chair during the procedure. The real stimulation was applied above the left dorsolateral prefrontal cortex, targeted by guiding the coil to the position F3 according to the international 10–20 system for electroencephalography electrode placement (Herwig et al, 2003b). The real stimulation intensity was determined as 110% of the individual resting motor threshold (Rossini et al, 1994). Inter-individual differences in cortical excitability and the use of different stimulators were thereby taken into account. Stimulations were performed with a frequency of 10 Hz, trains of 2 s, inter-train-intervals of 8 s, 100 trains per session, 2000 stimuli per day on 15 subsequent working days. Sham stimulation was applied 5 cm lateral to F3, perpendicular to the parasagittal plane, above the left temporal muscle; in this position the coil–cortex distance is essentially larger (more than 3 cm ν. 1–1.5 cm) than at F3, and the electromagnetic field reaching the cortex was therefore substantially weaker. To further reduce the possible effectiveness of the sham stimulation the coil was angled at 45°, touching the skull not with the centre but with the rim opposite the handle, and the stimulation intensity was reduced to 90% of motor threshold. Although the angling of the coil might have been registered by the patients as being different from the coil handling involved in measuring the motor threshold, this was a compromise made in an attempt to make the sham condition as similar as possible concerning side-effects to the real one but with minimum efficacy. Owing to the substantially weaker electromagnetic field reaching the cortex in this condition compared with real rTMS, neuronal depolarisation (Loo et al, 2000; Lisanby et al, 2001) was unlikely, as was any possible antidepressant effect. Nevertheless, this form of sham stimulation had the effect of inducing local sensations above the temporal muscle similar to the disturbances caused by the real stimulation (Praeg et al, 2005), helping to reduce bias from patient awareness of the difference between the two applications (Abler et al, 2005). Using a sham coil with no stimulation would have been even more different from real stimulation because of the absence of local sensations compared with the experience of motor threshold determination.

**Concomitant treatments**

In order to integrate rTMS in a naturalistic routine clinical setting, and for ethical and safety reasons, rTMS was applied in parallel with a standardised antidepressant medication or as monotherapy when no medication was possible. The stimulation sessions were started together with a venlafaxine or mirtazapine treatment, both selected because of their combined serotonergic and noradrenergic profile in order to rule out neurotransmitter-specific confounding effects. Prior antidepressant medication was washed out (4 weeks). Venlafaxine was started at a dosage of 75 mg per day in the first week, and mirtazapine at a dosage of 15 mg per day. Both treatments could be increased later according to clinical need as evaluated by the responsible psychiatrist. No other antidepressant or concomitant antipsychotic medication was allowed. A maximum of 1.5 mg lorazepam per day was permitted as crisis medication.

Patients whose condition had been stable on lithium treatment for at least 3 months before starting rTMS were allowed to continue taking this medication. Anticonvulsants were not allowed. Non-psychiatric medication was continued as needed and documented. All other treatments, such as psychotherapy and supportive therapies (music, occupational therapy, etc.), were also continued and documented, and compared between the real and the sham stimulation group.

**Efficacy variables and statistical procedure**

Baseline values were analysed with descriptive statistics. Frequencies were calculated for categorical data and means and standard deviations for quantitative variables. Furthermore, the baseline values of the real and the sham groups were compared with chi-squared tests for categorical variables or t-tests for quantitative variables.

The primary objective was to demonstrate that rTMS adjunctive to standard antidepressant treatment results in a greater number of responders (defined as patients with an improvement in scores on at least two of the three rating scales by at least 50% after 3 weeks of rTMS) than sham treatment (primary hypothesis). The secondary objective was to show a greater decrease in the depression rating scores with real rTMS than by sham treatment (secondary hypothesis). Remission was defined descriptively as a score of 10 points or below in all three scales. The BDI, HRSD and MADRS rating scales were administered prior to the stimulation sessions (rating 1); after 1 week and 2 weeks (ratings 2 and 3); at the end of the stimulation series after 3 weeks (rating 4); and at a follow-up interview 3 weeks later (rating 5). The first rating was made on the day before the stimulation period commenced. If rTMS was started the day after recruitment, the recruitment ratings were considered instead.

On the basis of previous reports (e.g. Pascual-Leone et al, 1996; George et al, 2000; Padberg et al, 2002; Herwig et al, 2003a) and assuming a clinically meaningful response in the real treatment group, we assumed a response rate of 50% due to augmentative and accelerative effects of rTMS after 3 weeks of stimulation compared with a sham response rate of 20% with the response due to medication assumed to occur later. Accordingly, the calculation of the sample size indicated that 45 patients were needed in each group to detect a difference in response rates between groups with 80% power at a 5% significance level. Presuming an estimated withdrawal rate of 20%, we aimed to include 120 patients in the study.

The primary efficacy variable analysed in the intention-to-treat set was treatment response. The comparison between treatment groups was performed by means of a Wald chi-squared test in a logistic regression model for the primary efficacy variable, adjusting for the stratification variables ‘centre’ (the centres Munich, Regensburg and Vienna, which had a joint rTMS training, were pooled in order to avoid numerical problems due to too small
sample sizes), and ‘HRSD’ (score ≤ 30 vs. > 30). Treatment × centre and treatment × HRSD interactions were tested in the model but were eliminated because P values exceeded 0.05. Results are described using odds ratios, 95% confidence intervals and P values. Secondary efficacy variables were the absolute and relative changes from rating 1 to 4 and 5 (before and after 3 weeks of stimulation, and at the follow-up) in the depression scores on HRSD, MADRS and BDI. They were compared between treatment groups using an F-test in a three-way analysis of variance (ANOVA) with treatment, centre and HRSD score as the main effects. Treatment × centre and treatment × HRSD interactions were again tested, and eliminated as P values were greater than 0.05. Least square means with 95% confidence intervals and P values for the comparisons between groups are reported.

Additional explorative analyses assessing the interaction effect of age (≤ 60 years vs. > 60 years), gender, device type and concomitant medication with treatment on the primary end-point were performed, by also including age or gender respectively in the models used for efficacy analyses. Owing to associations between device type and centre, device type was used instead of centre in the respective models.

All statistical analyses were performed with the Statistical Analysis System software package, version 8.02 for Windows.

**RESULTS**

**Participants**

The intention-to-treat (ITT) sample comprised 127 patients (Fig. 1, Table 1). The study commenced in 2003, and most of the patients were recruited between June 2004 and November 2005 after the researchers obtained a supporting grant from the German Research Foundation. The numbers of patients recruited by the different centres were Ulm n=37, Würzburg n=24, Rostock n=21, Tübingen n=16, Regensburg n=14, Munich n=11 and Vienna n=4. Of the 127 patients, 62 were randomised to the real stimulation group and 65 to the sham group. In the period between enrolment and start of the treatment, 5 patients showed an improvement in their depressive symptoms such that they no longer fulfilled the inclusion criteria. Two patients had to be excluded during or after stimulation (1 because of psychotic symptoms and 1 because a data review revealed an erroneous inclusion) and were not considered for the per protocol analysis. Fifteen participants withdrew during the stimulation series, 6 from the real intervention group and 9 from the sham group, for the reasons detailed in Fig. 1. Thus, 105 patients received stimulation according to protocol over the whole period of 3 weeks: 52 with real stimulation and 53 with the sham condition. Follow-up ratings 3 weeks after the end of the stimulation sessions were performed in 50 participants in the real group and 48 in the sham group; the remaining patients refused to participate or could not be contacted. Administration of concomitant medication was similar in both groups, including mean dosages (Table 1). Treatment groups were similar with respect to a continuation of supportive treatments such as occupational therapy, music therapy, relaxation techniques, supportive psychotherapy (real, n=52; sham, n=49) and, if established, a continuation of cognitive–behavioural or interpersonal therapy (real, n=19; sham, n=20). In the frame of the multiple comparisons of the baseline characteristics we found the real

---

**Fig. 1** CONSORT flowchart (rTMS, repetitive transcranial magnetic stimulation).
group to include more women than the sham group and to score marginally higher on the HRSD in the ITT set (no difference in an additional testing of the PP set; \(P=0.08\)), but not on the BDI or the MADRS. The other features and clinical baseline characteristics were similar in the two groups (Table 1).

### Table 1  Baseline characteristics of the real and sham intervention groups (n=127)

<table>
<thead>
<tr>
<th></th>
<th>Real n=62</th>
<th>Sham n=65</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender: male/female, n/n</td>
<td>18/44</td>
<td>33/32</td>
<td>0.02</td>
</tr>
<tr>
<td>Age, years: mean (s.d.)</td>
<td>50 (15)</td>
<td>49 (13)</td>
<td>0.73</td>
</tr>
<tr>
<td>Age (&lt;=60/&gt;&gt;60, n/n</td>
<td>43/19</td>
<td>49/16</td>
<td></td>
</tr>
<tr>
<td>Years of education: mean</td>
<td>11 (4)</td>
<td>12 (4)</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Depressive episode

- Unipolar/bipolar depression, n/n: 60/2 vs. 59/6, \(P=0.27\)
- Somatic syndrome, n: 14 vs. 18, \(P=0.55\)
- Age at onset, years: mean (s.d.): 38 (16) vs. 38 (16), \(P=0.99\)

Depression scale scores prior to intervention

- BDI score: mean (s.d.): 26.8 (8.9) vs. 27.0 (10.3), \(P=0.87\)
- MADRS: mean (s.d.): 28.0 (7.0) vs. 27.1 (6.3), \(P=0.30\)

Outcome

- Number of episodes (including current): 1-3 vs. at least 4, n/n: 25/36 vs. 26/36, \(P=0.76\)
- Treatment resistance present, \(n^2\): 9 vs. 10, \(P=0.92\)
- Psychosocial burden present, \(n\): 33 vs. 30, \(P=0.48\)
- Family history of depression, n: 20 vs. 31, \(P=0.10\)
- Medication: Venlafaxine, Patients, n: 30 vs. 28, \(P=0.86\)
- Mirtazapine, Patients, n: 28 vs. 31
- Supportive therapy (music, etc.), n: 52 vs. 49

Other therapies

- Specific psychotherapy, n: 19 vs. 20, \(P=0.99\)
- Supportive therapy (music, etc.), n: 52 vs. 49, \(P=0.23\)

**BDI, Beck Depression Inventory; MADRS Hamilton Rating Scale for Depression; MADRS Montgomery–Åsberg Depression Rating Scale.**

1. Values of chi-squared and t-tests as appropriate.
2. No response to two different antidepressant medication trials and one combination treatment.

Primary and secondary efficacy outcome

Within the ITT sample the analysis of treatment response revealed 19 responders (31%) in the real condition and 20 responders (31%) in the sham condition vs. 33 non-responders (53%) and 33 non-responders (51%) respectively. The remaining patients withdrew from the trial or were excluded (real, \(n=10, 16\%\); sham, \(n=12, 18\%\); Fig. 1). For the ITT analysis of primary efficacy, missing values for the patients who withdrew were recorded as non-response. After adjusting for centre and HRSD score at the start of the study, there was no significant difference in response rates between the different groups (OR=1.0, 95% CI 0.5-2.2, Wald \(\chi^2\) test, \(P=0.962\); Table 2). There was no meaningful difference in the response rates between the centres (\(P=0.339\)).

The ANOVA of the secondary efficacy variables, i.e. the absolute and relative changes from rating 1 to rating 4 (end of the rTMS period; Table 2, Fig. 2) and rating 5 (follow-up) of the depression scores on the HRSD, MADRS and BDI, revealed no difference between the real and sham groups at the end of the stimulation sessions. In the per protocol data-set, logistic regression showed no difference in the responder rates between the real and sham stimulation groups at any point during the course of stimulation, and thus no accelerated antidepressant effect (Fig. 3). Further, there was no meaningful difference in the responder rates between the treatment groups after the follow-up period (Wald \(\chi^2\) test, \(P=0.34\)). With regard to the absolute and relative changes in the rating scores, no meaningful difference was observed between the real and sham stimulation groups in the ratings after 1 week, after 2 weeks and at follow-up (Fig. 4). Remission of depression was found in 6 people in the real group and 10 people in the sham group.

Explorative analyses did not show any meaningful interaction effect of age, gender, device type or concomitant medication with treatment on the primary efficacy outcome.

### Side-effects

Patients complained of the following side-effects related to rTMS: headache (real, \(n=3\); sham, \(n=1\)), dizziness (real, \(n=0\); sham, \(n=1\)), painful local sensation (real, \(n=1\); sham, \(n=2\)) and nausea (real, \(n=1\); sham, \(n=0\)). Most patients reported that the stimulation generally caused an uncomfortable local sensation but they did
The aim of this multicentre trial was to observe no epileptic seizure or other side-effect. We did not complain about this as a side-effect. We observed no epileptic seizure or other severe side-effect.

DISCUSSION

The aim of this multicentre trial was to investigate the antidepressant effect of rTMS as an augmentative and/or accelerating treatment to simultaneously initiated antidepressant medication in a routine clinical setting. We did not find beneficial effects of active rTMS compared with the sham condition with regard to responder rates or changes in the rating scores. Furthermore, no acceleration of a clinical improvement was observed. No severe side-effect such as epileptic seizure occurred, indicating that the method may be considered to be safe within the frame of our study design and as far as the limits of our sample size allow.

Transcranial magnetic stimulation depolarises neurons in targeted cortex areas focally and non-invasively through induction of a transient electromagnetic field that is generated by a pulsed electrical current running through a wound copper coil. The induction of local and trans-synaptically mediated metabolic and biochemical changes in pathophysiology relevant brain areas was suggested as a rationale for an antidepressant effect (Post & Keck, 2001). The left dorsolateral prefrontal cortex was selected as a main target area for stimulation in patients with depression on the basis of imaging studies that attributed depressive symptoms to a regional hypometabolism which might be upregulated by rTMS (Pascual-Leone et al., 1996). The antidepressant properties of rTMS have now been investigated for more than 10 years, and initial positive studies elicited hope in both the scientific community and the public. Presumably in routine clinical care rTMS would be mainly applied concomitantly with other antidepressant treatments; for this reason an additional benefit of rTMS should be demonstrated in controlled clinical trials.

Table 2 Analysis of efficacy

<table>
<thead>
<tr>
<th></th>
<th>Real (n=62)</th>
<th>Sham (n=65)</th>
<th>OR (95% CI)</th>
<th>Difference in LSM (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention-to-treat sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>62</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response after 3 weeks</td>
<td>19 (31)</td>
<td>20 (31)</td>
<td>1.0 (0.5 to 2.2)</td>
<td>0.962</td>
<td></td>
</tr>
<tr>
<td>Per protocol sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>52</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response after 3 weeks</td>
<td>19 (37)</td>
<td>20 (38)</td>
<td>1.0 (0.4 to 2.1)</td>
<td>0.906</td>
<td></td>
</tr>
<tr>
<td><strong>Secondary analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per protocol sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>52</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression scale ratings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>after 3 weeks of rTMS, mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute change in BDI</td>
<td>11.3 (9.2)</td>
<td>9.4 (9.6)</td>
<td>1.18 (1.18 to 5.5)</td>
<td>0.324</td>
<td></td>
</tr>
<tr>
<td>Absolute change in HRSD</td>
<td>10.5 (6.2)</td>
<td>8.7 (8.0)</td>
<td>1.17 (1.0 to 4.4)</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td>Absolute change in MADRS</td>
<td>11.1 (7.5)</td>
<td>10.8 (9.4)</td>
<td>0.1 (2.9 to 3.2)</td>
<td>0.927</td>
<td></td>
</tr>
<tr>
<td>Relative change in BDI, %</td>
<td>39.3 (30.7)</td>
<td>32.4 (38.0)</td>
<td>6.7 (6.8 to 20.3)</td>
<td>0.328</td>
<td></td>
</tr>
<tr>
<td>Relative change in HRSD, %</td>
<td>43.0 (24.9)</td>
<td>38.2 (34.0)</td>
<td>4.6 (6.6 to 15.9)</td>
<td>0.417</td>
<td></td>
</tr>
<tr>
<td>Relative change in MADRS, %</td>
<td>38.4 (27.0)</td>
<td>38.5 (32.9)</td>
<td>0.5 (11.4 to 10.4)</td>
<td>0.927</td>
<td></td>
</tr>
</tbody>
</table>

BDI, Beck Depression Inventory; HRSD, Hamilton Rating Scale for Depression; LSM, least square mean; MADRS, Montgomery–Asberg Depression Rating Scale.

Comparison with other rTMS treatment trials and limitations

Our multicentre results are in contrast to several positive reports from single-centre studies of rTMS for depression (reviewed by Burt et al., 2002; Martin et al., 2003; Loo & Mitchell, 2005), but they are in line with other negative reports (Loo et al., 2003; Nahas et al., 2003; Poulet et al., 2004; Minissi et al., 2005). Our results are to be compared in particular with studies addressing the specific issue of rTMS as an add-on or augmentative treatment to antidepressant medication. Recent studies of this topic that reported positive
results require discussion in more detail in relation to our results. In a trial investigating rTMS (5 Hz, 120% of motor threshold, 1200 stimuli per day) above the left dorsolateral prefrontal cortex, given in parallel with amitriptyline titrated up to a therapeutic dosage during the week before starting rTMS, beneficial effects were found already after the first week of stimulation and were sustained for the stimulation period of 4 weeks (Rumi et al., 2005). Another study, combining rTMS (15 Hz, 100% of motor threshold, 900 stimuli per day) above the left dorsolateral prefrontal cortex with venlafaxine, citalopram or sertraline started simultaneously and titrated up quickly, found beneficial effects after 2 weeks of stimulation, but these benefits had disappeared at the follow-up assessment 3 weeks later (Rossini et al., 2005). Concerning stimulation parameters, the values used in our study (10 Hz, 110% motor threshold) were between those of the two studies mentioned above but our daily amount of stimuli was higher, so that these differences can hardly account for our negative results. A further study reporting beneficial effects (Anderson et al., 2007) applied rTMS at 10 Hz, 110% of motor threshold, 1000 stimuli per day, three times per week for 4–6 weeks while the patients were maintained on established medication. Here, the difference from our results might be due to unchanged medication in largely treatment-resistant patients, with thus no further medication effect as indicated by a low response in the sham group (7%), and to the longer stimulation period. Generally, different regimens of co-medication in these studies are to be considered when comparing the results. Other add-on rTMS studies with negative results might have suffered from insufficient stimulation parameters such as sub-threshold intensity and low number of stimuli (Poulet et al., 2004).

The stimulation parameters for our study were chosen as those most likely to have a possible antidepressant effect, based on the evidence available at the time of study conception: higher intensities (≥100% of motor threshold), frequencies (≥5 Hz) and total amounts of stimuli (≥10,000); treatment periods of at least 10 days; and targeting the left dorsolateral prefrontal cortex (e.g., Pascual-Leone et al., 1996; Padberg et al., 2002; Grunhaus et al., 2003; Herwig et al., 2003b; Loo et al., 2003; Martin et al., 2003). One might argue that our chosen stimulation period of 3 weeks was too short. However, the above-mentioned papers and the majority of other relevant studies reported positive effects even earlier, i.e. after 1–2 weeks of stimulation. A single-centre study that used the same parameters concerning intensity, frequency, location and duration as we did, albeit with fewer stimuli per day (1600) in 5 s trains and with a different study design, recently reported beneficial rTMS effects in treatment-resistant depression (Avery et al., 2006). Thus, on the basis of the literature, we could have expected to detect an antidepressant effect from the stimulation parameters used in our trial. The improvement observed in both groups of our study may be explained as an effect of medication, a general placebo effect or the spontaneous course of the disease. Further, clinical factors such as short episode duration and lack of treatment resistance, whenever we had a strict definition, in some of our patients might have accounted for the generally good antidepressant response. Accordingly, one may argue that a possible antidepressant effect of rTMS might have been hidden by the medication effect and by these clinical factors; but one can at least state that no beneficial effect of rTMS in addition to newly initiated medication with mirtazapine or venlafaxine at the standard lower dose range was observed. In this context, it may also be argued that our study might have been underpowered and that more patients should have been included in order to reveal a significant difference. However, we observed the same rates of responders (31%) in both groups, implying that even if many more patients had been treated the outcome in the primary efficacy variable would not have been any different. As concerns the number of included patients, it should be noted that this study is one of the largest of rTMS in depression reported to date. The antidepressant response found in our study for both stimulation conditions is comparable with the results reported for a 3-week period of treatment (within longer courses) in pharmacological studies that investigated the antidepressant response on mirtazapine and venlafaxine in terms of changes in HRSD and/or MADRS rating scores and response rate (e.g., Amini et al., 2005; Shelton et al., 2006). Accordingly, we found no evidence that the response to rTMS and medication in our study was superior to that reported by studies that investigated solely medication effects. Concerning patient characteristics, we found no influence of age and gender on outcome. Although other studies suggested an age-dependent rTMS effect with less efficacy in the elderly (Mosimann et al., 2004; however, that study used lower intense stimulation parameters), in our study neither the younger nor the older patients responded to rTMS. Further, considering our gender distribution, gender showed no effect on treatment outcome in the explorative analysis, which also would not have been supported by any evidence in the literature. The HRSD baseline scores were slightly higher in the real stimulation group, whereas MADRS and BDI scores did not show any difference between the groups. Within a set of multiple comparisons it was likely that differences would be observed in relation to distinct features. The mean absolute difference in HRSD scores, however, was less than 2 points and therefore clinically marginal. Further, the analyses had been adjusted for HRSD score (≤30 v. >30) at the start of the study, and no different outcome dependent on HRSD score was observed. Considering these facts and that the study outcome was negative, there was no meaningful bias in our view. We further found no influence of stimulator type or concomitant medication on treatment outcome, and no difference in the clinical baseline variables.

Meta-analyses addressing rTMS studies in depression draw critical conclusions concerning the applied methodology and the clinical significance of the results. Kozel & George (2002) found a mean difference in improvement in studies using real vs. sham rTMS of 3 points on the HRSD, the clinical impact of which appeared to be marginal. Furthermore, for methodological reasons they considered only a small number of the studies on this topic. Martin et al. (2003) also criticised methodological issues and concluded that there was no strong evidence of benefit from using rTMS to treat depression, although the small sample sizes of the studies did not allow the possibility of such an effect being excluded. A recent meta-analysis concluded that rTMS may not differ from sham treatment in major depression (Couturier, 2005). However, that analysis also excluded several studies because of methodological issues and therefore based its outcome on only a few studies. Therefore, the current literature and our data dampen early expectations about positive effects of rTMS on depression and indicate that one should be careful about generally implementing rTMS in clinical practice.
**REFERENCES**


prefrontal transcranial magnetic stimulation for the treatment of resistant major depression. Psychological Medicine, 33, 33–40.


World Health Organization (1992) The ICD–10 Classification of Mental and Behavioural Disorders. WHO.
Intergenerational transmission of health beliefs in somatoform disorders

Exploratory study

TAMSIN MARSHALL, DAVID P. H. JONES, PAUL G. RAMCHANDANI, ALAN STEIN and CHRISTOPHER BASS

Summary  Children of parents with a range of psychiatric disorders are at increased risk of developing psychological disturbance themselves. There is growing evidence that this includes children who have parents with a chronic somatoform disorder. The health beliefs of children with a parent with a somatoform disorder were compared with those of children with a parent with an organic physical disorder. Children of parents with somatoform disorder scored higher on bodily preoccupation and disease phobia scales and their health beliefs showed similarities to the beliefs of their parents.

Declaration of interest  None.

It is well established that parental mental illness is associated with an increased risk of psychological problems in children (Barnes & Stein, 2003). There have been few studies of the children of parents with a somatoform disorder, despite the fact that such disorders are common and account for a significant proportion of healthcare utilisation (Barsky et al., 2001). There is some evidence for the clustering of somatisation in families (Garralda, 2000), indicating possible intergenerational transmission and raising questions about the mechanism of any such transmission. This study aimed to examine whether the children of parents with a somatoform disorder had more abnor-}

child in the age range 8–16 years. They were identified from a database of all patients referred to a liaison psychiatry service at a large teaching hospital.

The comparison group comprised parents with a chronic organic physical illness of at least 5 years’ duration and their oldest child in the same age range as the study group. They were recruited from hospital out-patient clinics for patients with inflammatory bowel disease, rheumatoid arthritis and multiple sclerosis. These disorders were chosen to ensure that the comparison group had chronic illnesses that were punctuated by recurrences of severe ill health and were therefore similar in pattern of illness to somatoform disorders. It was also necessary to select illnesses that affect people at an age when they were likely to be parents of school-age children. A mix of conditions was used to recruit a heterogeneous comparison group, as the patients with somatoform disorder also had varied presentations and a range of symptoms.

All parents and children of secondary school age gave written consent. Verbal assent was obtained from the younger children. The local research and ethics committee approved the research protocol.

The Illness Attitudes Scale (IAS; Kellner et al., 1987) was used to measure the parents’ health beliefs. It consists of nine scales: worry about illness; concern about pain; health habits; hypochondriacal beliefs; thanatophobia (fear of death); disease phobia; bodily preoccupations; treatment experience; and effects of treatment. It has a 5-point response scale for each question.

A modified and validated version of the IAS was used to measure the health beliefs of the children aged 11 and over. The modified IAS was adapted by Eminson et al. (1996) from the original IAS for use with 11- to 16-year-olds. For those children under the age of 11 years, the questionnaire was modified further, by simplifying the language and omitting some questions on smoking and alcohol.

METHOD

The study group consisted of parents with a DSM–IV diagnosis (American Psychiatric Association, 1994) of somatoform disorder of at least 5 years’ duration and their oldest hospital out-patient clinics were identified as meeting criteria for the comparison group. Of these, 15 (68%) agreed to take part.

The groups were comparable with respect to age of child and parent, gender of child and parent, ethnicity, marital status, social class and whether the studied parent was the child’s main carer.

Children in the somatoform group scored significantly higher overall on the IAS and on the following sub-scales: bodily preoccupations; disease phobia; treatment experience; and effect of treatment (Table 1). A similar pattern of differences was seen with parental scores, with parents in the somatoform group scoring significantly higher on the following IAS sub-scales: bodily preoccupation; hypochondriacal beliefs; treatment experience; and effect of treatment (Table 1).

RESULTS

Thirty-three parents with a diagnosis of somatoform disorder were identified as meeting the study criteria. Of these, three were excluded because they had moved or were not contactable. Of the 30 who were approached to take part in the study, 18 (60%) agreed to participate. Twenty-two patients currently attending three separate hospital out-patient clinics were identified as meeting criteria for the comparison group. Of these, 15 (68%) agreed to take part.

The groups were comparable with respect to age of child and parent, gender of child and parent, ethnicity, marital status, social class and whether the studied parent was the child’s main carer.

Children in the somatoform group scored significantly higher overall on the IAS and on the following sub-scales: bodily preoccupations; disease phobia; treatment experience; and effect of treatment (Table 1). A similar pattern of differences was seen with parental scores, with parents in the somatoform group scoring significantly higher on the following IAS sub-scales: bodily preoccupation; hypochondriacal beliefs; treatment experience; and effect of treatment (Table 1).

DISCUSSION

To our knowledge, this is the first study to show that the health beliefs of children who have a parent with a somatoform disorder are different from those of children whose parents have an organic physical condition. Children with a parent with a somatoform disorder had higher scores on a measure of problematic health cognitions; in particular they reported more bodily preoccupation and disease phobia. Furthermore, the different health beliefs in the two groups of children showed similarities to the different health beliefs in the parents. The
inclusion of a group of children whose parents had an organic illness meant that both groups were likely to be similar in terms of experiences related to having an ill parent.

The main limitation of this exploratory study is the sample size. The findings are therefore preliminary and require replication. It might have been easier to recruit a larger community sample of parents with less severe somatising problems but at the risk of seeing a smaller effect size. However, despite the small sample size, the study shows significant differences in some specific health cognitions between the two groups of children. The finding that these differences were similar to those seen in the two groups of parents indicates that there may be intergenerational transmission of problematic cognitions and raises interesting questions about the mechanisms by which such transmission may occur.

The mechanisms for intergenerational transmission are likely to be multiple and include genetics, parental psychopathology, family stresses and parenting style (Walker, 1999; Crane & Martin, 2004), as well as interactions between these factors. It has been hypothesised that mechanisms of transmission in somatisation may include maternal modelling and reinforcement of illness behaviours (Walker et al., 1993; Craig et al., 2002; Crane & Martin, 2004).

Somatoform disorders are important in terms of prevalence, levels of suffering and cost to health services (Barsky et al., 2001). The current study suggests that when these patients are parents their beliefs and behaviours have an impact on their children. By understanding more about the mechanisms of intergenerational transmission of health beliefs, it might be possible to develop effective interventions aimed at preventing the development of somatoform disorders.

**ACKNOWLEDGEMENTS**

We thank the parents and children who took part in the study and our colleagues in the departments of rheumatology, neurology and gastroenterology. We thank Annie Shrier for data entry. P.R. and A.S. are supported by the Wellcome Trust, which also supported this study. The study won the prize for best poster at the 2005 Annual Residential Meeting of the Royal College of Psychiatrists’ Liaison Psychiatry Faculty.

**REFERENCES**


**Table 1** Health beliefs of parents and children as measured by the Illness Attitudes Scale

<table>
<thead>
<tr>
<th>Illness Attitudes Scale sub-scale</th>
<th>Parents</th>
<th>Children</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Somatoforn disorder (n=18)</td>
<td>Organic disorder (n=15)</td>
</tr>
<tr>
<td>Worry about illness</td>
<td>Median (interquartile range)</td>
<td>6 (3–8.25)</td>
</tr>
<tr>
<td>Concern about pain</td>
<td>4 (0.75–6.25)</td>
<td>4 (2–6)</td>
</tr>
<tr>
<td>Hypochondriacal beliefs</td>
<td>2 (0.75–5)</td>
<td>0 (0–1)</td>
</tr>
<tr>
<td>Health habits</td>
<td>6 (3–9)</td>
<td>8 (5–10)</td>
</tr>
<tr>
<td>Thanatophobia</td>
<td>3.5 (0–5.25)</td>
<td>1 (0–4)</td>
</tr>
<tr>
<td>Disease phobia</td>
<td>1.5 (0–4.25)</td>
<td>1 (0–3)</td>
</tr>
<tr>
<td>Bodily preoccupations</td>
<td>3 (1.75–4.25)</td>
<td>1 (0–4)</td>
</tr>
<tr>
<td>Treatment experience</td>
<td>8.5 (7–10)</td>
<td>6 (5–8)</td>
</tr>
<tr>
<td>Effect of symptoms</td>
<td>8.5 (7.5–10)</td>
<td>4 (3–8)</td>
</tr>
<tr>
<td>Overall score</td>
<td>42 (32.75–56.25)</td>
<td>32 (30–46)</td>
</tr>
</tbody>
</table>


Effects of training community staff in interventions for substance misuse in dual diagnosis patients with psychosis (COMO study)

Cluster randomised trial


Summary  A cluster randomised controlled trial was used to investigate the effectiveness of training staff in 13 London community mental health teams (CMHTs) to deliver substance misuse interventions to patients with psychosis and comorbid substance misuse (‘dual diagnosis’). The primary hypotheses, which were that experimental group patients would spend fewer days in hospital over 18 months of follow-up and show reduced alcohol and drug consumption, were not confirmed, although confidence intervals were wide for some outcomes. Current UK policy guidance advocates training CMHT professionals to deliver dual diagnosis interventions, but the effectiveness of this strategy has not so far been demonstrated.

Declaration of interest  None.

The adverse outcomes associated with psychosis and comorbid substance misuse (‘dual diagnosis’) have been well documented, but treatments supported by substantial evidence are few (Jefery et al., 2000; Tyrer & Weaver, 2004). Interventions that show some benefit have tended to involve relatively intensive treatment of selected populations by specialist therapists (Barrowclough et al., 2001; Bellack et al., 2006); generalisability to routine settings is unknown. However, commentaries and policy guidance on dual diagnosis in England have favoured ‘mainstreaming’ dual diagnosis interventions by integrating them into care provided by existing clinical teams (Johnson, 1997; Weaver et al., 1999; Department of Health, 2002). The effectiveness of this strategy is untested.

Our aim was to investigate whether a training and supervision intervention delivered to community mental health team (CMHT) case managers would improve patient outcomes. At the patient level, the primary hypotheses were that, compared with controls, patients on the case-loads of experimental group case managers would make less use of in-patient services and would be consuming smaller quantities of substances when assessed 18 months later.

METHOD

A cluster randomised controlled trial design was employed, each cluster consisting of the patients on a particular staff member’s case-load.

All permanent case managers in 13 London CMHTs were invited to participate. Their case-loads were screened for patients who met study criteria for dual diagnosis, and all who did were included in the sample. This screening stage involved first identifying patients with clinical diagnoses of schizophrenia, another non-affective functional psychosis, or bipolar affective disorder. With guidance from researchers, case managers rated each of these patients using the Clinician Alcohol and Drug Use Scales (Drake et al., 1996). Patients identified as misusing or dependent on at least one substance met our study criteria for dual diagnosis. Case managers were randomised to intervention or control groups by an independent statistician. All patients identified as eligible for the trial entered the experimental or control group according to their case manager’s assignment.

The experimental intervention consisted of a treatment manual, a 5-day training course in assessment and management of dual diagnosis, and subsequent monthly supervision. Motivational interviewing was a central source (Swanson et al., 1999), and the training also drew on cognitive–behavioural relapse prevention techniques (Irvin et al., 1999). The control group received CMHT management as usual with no specific dual diagnosis intervention. To reduce contamination, experimental group staff were asked to avoid sharing manuals and details of training.

At baseline, socio-demographic and clinical details of all patients were recorded. At baseline and after 18 months, data were collected on the two primary outcome measures: (a) hospital bed use over the preceding 18 months, recorded using best available information from patient interview, clinical records and local electronic patient data systems; (b) substance use over the preceding month, documented at patient interview using the Maudsley Addictions Profile (Marsden et al., 1998).

Secondary outcomes relating to adverse events, symptoms and social functioning and staff-level outcomes were also assessed, but are not reported here. Interviews with patients were carried out whenever possible; for patients who were not available, ethical approval was obtained to gather data from staff on their characteristics and the bed use outcome.

RESULTS

Seventy-nine case managers participated. Of the 1560 patients on their case-loads, 232 met criteria for dual diagnosis. Forty of the 79 case managers were randomised to the experimental group and 39 to the control group. This yielded 127 patients with dual diagnosis on case-loads of case managers in the experimental group and 105 on control group case-loads. Miles et al. (2003) have described the characteristics of the sample. Experimental and control groups were similar except for an imbalance for White ethnic group (61% of the control group v. 43% of the intervention group). CONSORT diagrams of staff and patient flows through the study are given in data supplement 1 to the online version of this paper.

Three patients died during the 18-month follow-up period. Of the remaining 229 patients, 77 (62%) of the intervention group and 77 of the control group (74%) were interviewed at follow-up (P=0.079). Bed use data were obtained for 113 intervention and 97 control group members. We defined experimental group patients as having received the intervention as intended if their case managers had attended at least 4 days of training and if they had remained on the case-load of a trained case manager for at least 9 months: just 45 of the 127 experimental group patients met these criteria. Eighty-six of the 105 control group members fitted the study definition of
having remained in their intended treatment group, which required them to have remained on a CMHT case-load for at least 9 months and not to have been taken on by a trained case manager.

Details of outcomes are shown in data supplement 2 to the online version of this paper. For bed use, there was no evidence of a difference between experimental and control groups (mean bed use for experimental group: 74.9 days (s.d.=142.6) over 18 months follow-up; for control group 71.8 days (s.d.=128.1), P=0.30 following log transformation and adjustment for baseline). However, standard deviations were higher than anticipated when carrying out the study power calculations, resulting in wide 95% confidence intervals. There was no significant difference in proportion of patients admitted during the follow-up period (43% of the experimental group v. 48% of the control group, P=0.18).

Self-reported alcohol and drug use of interviewed members of each group over the 30-day period before the follow-up interview are also shown in online data supplement 2. Neither the proportion who had consumed substances (74% of the experimental group and 71% of the control group for alcohol, 32% and 36% respectively for cannabis and 16% and 18% for other drugs) nor the quantity consumed over the month differed between groups. No difference in outcomes became significant after adjusting for baseline values.

**DISCUSSION**

The study’s strengths lie in external validity: the intervention took place in a routine National Health Service setting and was brief enough to be replicable in such settings, and all identified patients with dual diagnosis were included. Limitations include high attrition from the intended intervention, reliance on clinician substance misuse diagnosis and lack of masking. Fidelity was not measured, and we do not know to what extent case managers implemented the intervention as intended. Also, for the main outcomes, standard deviations were wider than anticipated when power was calculated: confidence intervals are thus wide and include the possibility of a substantial effect in either direction.

There was no evidence that the training intervention affected bed use or substance use. The limitations discussed above must be considered in interpreting this finding.

**REFERENCES**


Correspondence: Dr Sonia Johnson, Department of Mental Health Sciences, University College London, 2nd Floor, Charles Bell House, 67–73 Riding House Street, London W1W 7JE, UK. Email: s.johnson@ucl.ac.uk

(First received 29 January 2006, final revision 12 April 2007, accepted 19 April 2007)
Are people with schizophrenia more logical than healthy volunteers?

GARETH S. OWEN, JOHN CUTTING and ANTHONY S. DAVID

Summary  We tested a phenomenological hypothesis about theoretical and practical rationality in people with schizophrenia. This hypothesis states that in schizophrenia there is an enhancement of theoretical rationality. Our case-control experiment supported this hypothesis. Philosophical models of rationality that prioritise theoretical over practical rationality may thereby apply more in schizophrenic than in healthy states. The study is an example of how experimental psychopathology can illuminate areas of philosophical dispute that are difficult to settle by thought alone.

Declaration of interest  None.

There has been renewed interest in the phenomenological tradition in psychiatry (Owen & Harland, 2007) which offers a range of hypotheses about different psychopathological states. One notable hypothesis is that schizophrenia is an impairment of commonsense knowing (practical reason), with a preservation—or even accentuation—of systematic cognition (theoretical reason). This concept was made famous by the psychiatrist Minkowski (1927) and can even be found in the anthropological writings of Kant (1798) but has never been experimentally tested. Previous experimental work on rationality in schizophrenia has aimed to explain delusions in schizophrenia in terms of impairments of formal reasoning (Garety & Hemsley, 1995; Kemp et al, 1997). Results do not generally confirm this model (Cutting, 1997).

In line with the phenomenological hypothesis, we tested whether tasks that are correct from a theoretical (or formal logical) point of view but depart from practical knowledge (common sense) would be performed better by people with schizophrenia than by healthy controls.

METHOD

Most philosophers conceptualise theoretical rationality as formal logical rationality for which deductive logic is held as the paradigm. Practical rationality or ‘common sense’, however, has been more difficult to conceptualise. It is generally taken to denote non-formal rationality—a form of knowing that provides the background assumptions about the world that are the basis of human practice. It is tacit knowledge within a culture, and includes such things as the pre-theoretical knowledge that the sun rises in the east or that hospitals are buildings. Such knowledge is presupposed and used in everyday practice, and as such becomes something that is separate from theoretical knowledge. The concept of common sense is that there is a form of rationality that is independent of theoretical rationality. The experiment we report assumes these two concepts and takes them to be fundamentally different.

We operationalised theoretical reasoning using syllogisms that were deductively valid or invalid, and common sense using syllogistic content that strongly conformed to or departed from practical knowledge. Two types of syllogism were constructed, in each of which there was a conflict between deductive truth and commonsense truth. The first type was non-commonsense syllogisms that were valid (labelled ‘NCS’), for example:

- All buildings speak loudly;
- A hospital does not speak loudly;
- Therefore, a hospital is not a building.

The second type comprised commonsense syllogisms that were invalid (labelled ‘CS’), for example:

- If the sun rises, then the sun is in the east;
- The sun is in the east;
- Therefore, the sun rises.

Participants were asked by the interviewer (G.O.) to accept the first two sentences of each syllogism as true and then to decide on the truth or falsity of the third sentence. They were told that this rule applied to all the problems and were asked to state it repeatedly until it was clear that they understood it. All participants read the problems aloud. Syllogisms were scored as correct if they were answered logically.

To be more certain that our syllogisms did generate subjective conflict between a logical and a commonsense interpretation in healthy people, we had previously conducted an independent pilot study in which we tested 21 healthy individuals. Verbal reports confirmed the conflict between logical and commonsense interpretations. We discarded three syllogisms that accrued high scores on the basis that their common-sense content was too weak, leaving eight NCS syllogisms and seven CS syllogisms for inclusion in the study reported here.

Ethical approval for the study was gained and all participants gave informed consent. People diagnosed with schizophrenia using standardised criteria (DSM-IV; American Psychiatric Association, 1994) and healthy controls were asked to solve the syllogisms in a case–control design. Patients were selected from two inner-London psychiatric hospitals; the sources were two general in-patient wards and the out-patient and in-patient facilities of a single service specialising in schizophrenia. All participating patients were taking antipsychotic medication. The control group was selected from a wide variety of informal sources, including acquaintances, porters and staff at several hospitals, and advertisement. Exclusion criteria for both groups were age outside the range 18–65 years; premorbid IQ, estimated using the National Adult Reading Test (Nelson, 1994), outside the range 75–125 (as at extreme values this measure is a poor guide to full-scale IQ (Russell et al, 2000)); English not native language; other neurological or psychiatric disorder or substance misuse. Medical records were reviewed for all patients and a clinical interview was conducted by a psychiatrist (G.O.) to ensure that criteria were met. Of the 22 patients approached, two were excluded because of elicited histories of epilepsy or heavy substance misuse and three because of NART IQ score <75. Of the 21 potential control group members, one was excluded because aged >65 years and one because of IQ score >125.

Our primary measures were number of syllogisms correct as a total and as subsets according to type (NCS or CS). Potential confounding factors were considered to be IQ, age, gender and years of education.
All t-tests performed were two-tailed with equal variance not assumed. Using percentage logically correct as the dependent variable, we performed an inter-individual factorial analysis of variance testing for main effects by group (schizophrenia vs. control) and syllogism type (NCS vs. CS) and their interaction. Our hypothesis was that the schizophrenia group would outperform the control group.

RESULTS

Groups were well matched, with 17 patients and 19 controls. There was no significant difference between the groups in premorbid IQ ($t = -0.87, P = 0.4$), age ($t = 1.25, P = 0.22$) or years of education ($t = -0.06, P = 0.96$). About half (53%) of the control group were men, compared with 65% of the schizophrenia group.

Table 1 shows the group statistics. As predicted there was a highly significant main effect by group ($F_{1,36} = 8.002, P = 0.006$), with patients outperforming controls. There was also a main effect by syllogism type ($F_{5,180} = 52.916; P < 0.001$), but no interaction of syllogism type by group ($F_{5,180} = 0.157, P = 0.69)$. The main effect by syllogism type showed that both groups scored better on the NCS syllogism type than on the CS syllogism type. We take this to be the well-replicated ‘belief bias’ effect (Evans, 2002), i.e. that logic has a larger effect on unbelievable (NCS) than on believable (CS) conclusions.

In exploratory analysis of the group difference, the effect size using the Cohen’s $d$ statistic was 0.82 (large) for the CS syllogism type and 0.54 (medium) for the NCS syllogism type. Similarly, comparisons of means showed significance for the CS syllogism type ($t = -2.37, P = 0.026$) but not for the NCS type ($t = -1.65, P = 0.11$). This suggests that there might be an underlying interaction between syllogism type and group, with the CS syllogism type (commonsense reasoning) accounting for most of the group difference, and that our failure to find it was due to inadequate statistical power.

DISCUSSION

Our main results show that under conditions where common sense and logic conflict, people with schizophrenia reason more logically than healthy individuals. On a straightforward interpretation this is either because people with schizophrenia are better at logic or because they are worse at common sense. We present some exploratory evidence that it is because they are worse at common sense, but the question remains open.

A few limitations must be mentioned. The number of participants was small, experimental designs using philosophical concepts are novel and case-control studies cannot control for unknown confounding factors. For example, our stimuli did not allow for correct rejections of non-commonsense syllogisms or correct acceptance of commonsense syllogisms.

The results are intriguing because they shed light on reasoning in schizophrenia but also have significance beyond schizophrenia research. They suggest that in situations where commonsense knowledge is at stake, formal norms of rationality are violated by people with schizophrenia to a lesser extent than by healthy individuals. People with schizophrenia seem to have a bias towards theoretical rationality over and above practical rationality. It is an ongoing dispute within philosophy of science whether, as a matter of principle, theoretical reason has priority over practical reason or vice versa (Thagard, 2004). Given that schizophrenia is at its core a pathological state of thinking, our results suggest that concepts of rationality that prioritise theoretical reason over and above practical reason might apply more accurately in a pathological example of human thinking than in a healthy one. This is an example of how experimental psychopathology can shed light on fundamental philosophical debates that have not been settled by argument alone.

REFERENCES


Correspondence

EDITED BY KIRIAKOS XENITIDIS and COLIN CAMPBELL

Contents ■ Depression post-myocardial infarction ■ Substance misuse disguised as ADHD? ■ Heroin-assisted treatment: no difference in treatment retention ■ Factors in those who repeatedly self-harm ■ Attachment disorders: an evolutionary perspective

Depression post-myocardial infarction

Van Melle et al (2007) reported that cardiac prognosis post-myocardial infarction was not improved by antidepressant treatment (MIND-IT trial). The SADHART and ENRICHD trials reported similar findings and Carney & Freedland (2007), in their commentary in the same issue, suggest these negative findings are a result of insufficient statistical power in the trials. These results are disappointing but perhaps they might have been anticipated.

There is strong evidence that individuals with depression show increased morbidity and mortality from coronary heart disease (Rugulies, 2002) but the mechanisms involved remain unclear. Individuals with a history of recurrent depression, who are otherwise healthy, show increased inflammation, platelet activation, endothelial dysfunction, and reduced heart rate variability and baroreceptor sensitivity. However, with the exception of platelet function, which improves with selective serotonin reuptake inhibitors, these anomalies are not corrected by antidepressant treatment. Furthermore, endothelial function and baroreceptor sensitivity, which can lead respectively to progression of the atherosclerotic process and to sudden cardiac death, do not improve when depressive symptoms are in remission (Broadley et al, 2006). Thus there is no evidence that treatment of depressive symptoms post-myocardial infarction corrects these underlying pathological processes and, if it does not, cardiac outcomes disclosed by clinical trials are unlikely to show improvement irrespective of their statistical power. By analogy, although hyperglycaemia characterises diabetes, tight glucose control alone has only a modest impact on cardiovascular events. Similarly, depressive illness is characterised by acute episodes of depression, but other systemic abnormalities are present and persist between acute depressive episodes. Accordingly, it may be unreasonable to believe that treatments assessed by their influence on the affective state alone will reduce cardiovascular events.

Although it is important to alleviate the suffering associated with developing depression post-myocardial infarction and improve prognosis by addressing the secondary effects of depression (e.g. reduced adherence to treatment and poor health behaviours), treatment needs to be aimed at earlier stages of the disorder. Atherosclerosis begins in childhood and becomes manifest much later in life, with myocardial infarction as a very late presentation. Similarly, depression is a lifelong disorder with onset in early adulthood. It should be noted that currently depression is not even included in cardiovascular risk tables and that individuals with depression might benefit from introduction of statins, or other preventative measures.

We agree with Carney & Freedland (2007) that treatments for depression might alter the risk of cardiac events via pathways that are unrelated to their effects on depression. However, if the focus of research were shifted to the study of earlier stages of coronary heart disease in people with depression, this could be clarified by monitoring earlier indices of heart disease in relation to treatment of depression. It is also recognised that mechanisms for associations between depression and onset of heart disease may differ from those between depression and progression of coronary heart disease post-myocardial infarction. These pathways need to be better understood and present evidence suggests that survival times following myocardial infarction could be improved by developing treatments for depression that also target the underlying cardiovascular abnormalities and by augmenting these by preventative programmes for coronary heart disease in individuals with mood disorders.

Coronary heart disease and depression are two major public health problems and it is of concern that reports of treatments for depression failing to enhance survival post-myocardial infarction may result in less interest in studying the links between them.


A. Korszun Wolfson Institute of Preventive Medicine, Barts and the London, Queen Mary’s School of Medicine and Dentistry, London, UK. Email: a.korszun@qmul.ac.uk

S. Stansfeld, M. Frenneaux Wolfson Institute of Preventive Medicine, Barts and the London, Queen Mary’s School of Medicine and Dentistry, London, UK.
doi: 10.1192/bjp.191.5.455

Van Melle et al (2007) present findings from their randomised controlled trial examining the effects of antidepressant treatment for depression following myocardial infarction. I would like to comment on the design of the study. Patients were allocated to two arms: antidepressant treatment and care as usual. Patients in the care-as-usual arm were not told about their research diagnosis of depression. The authors quote Zelen (1979), thus implying that they are following the research design he proposed. However, Zelen’s method seems best suited to trials where there is a ‘gold standard’ control treatment available and the trial is attempting to evaluate a new experimental treatment (Zelen, 1979). In this design, the ethical concerns are mainly about randomising before consent is sought. It must be pointed out that after randomisation, consent is sought from patients in the experimental arm. If they decline, they are moved to the ‘gold standard’ arm (Torgerson, 2001). I am not sure whether the trial of van Melle et al fits into this category.

Furthermore, there are ethical issues about not informing patients about their diagnosis of depression. I am disappointed that the paper did not discuss these in further detail. Their information pack stated
that all patients were free to seek help for their mood problems. Patients may feel tired and low in mood but may not recognise this as depression, for which there are effective interventions available. Is it ethical to withhold information regarding the diagnosis from such patients? Will patients seek help if they are not told they have depression?

Performing research can raise difficult ethical issues and I hope this letter will encourage some debate on this.


A. Shetty  Crisis Team West, Queen’s Medical Centre, Derby Road, Nottinghamshire NG7 2UH, UK. Email: dradashshetty@yahoo.co.in
doi: 10.1192/bjp.191.5.455a

Van Melle et al (2007) found no difference in efficacy and cardiac prognosis between treatment with antidepressive medication and care as usual in patients with depression after myocardial infarction. Carney & Freedland (2007) commented that the lack of difference in efficacy prohibits the demonstration that effective treatment of depression improves survival. They emphasised the need for developing highly efficacious treatments for depression following myocardial infarction. Such a treatment, however, already exists, as electroconvulsive therapy (ECT), and has been shown to have superior efficacy compared with antidepressive medication (ECT UK Review Group, 2003).

A trial using ECT as an intervention will more likely find a superior efficacy compared with treatment as usual and may demonstrate that effective depression treatment improves survival. Because of concerns about the cardiac risks some textbooks do not recommend the use of ECT within 3 months of myocardial infarction. Zielinski et al (1993) found a higher rate of cardiac complications during ECT in patients with a pre-existing cardiac abnormality compared with patients with no pre-existing abnormality. Most complications, however, were transitory and did not prevent the completion of the ECT course. Rice et al (1994) found that ECT increased the risk of minor but not severe complications. They pointed to the advances in ECT techniques which have resulted in improved safety in cardiac patients. The risk of ECT has to be weighed against the risk of an inadequate treatment of depression, which is known to increase mortality (van Melle et al, 2007). Considering the high risk of cardiac events of 13% in the 18 months following myocardial infarction (van Melle et al, 2007), which may partly be attributable to the inadequate treatment of depression, treatment with ECT could be safer because of its superior efficacy as an antidepressant.


K. H. Kho  GGZ Delfland, St Jorisweg 2, 2612 GA, Delft, The Netherlands. Email: kkhkho@kk4all.nl
doi: 10.1192/bjp.191.5.456

Authors’ reply: To explain why anti-depressant treatment with selective serotonin reuptake inhibitors (SSRIs) does not improve cardiac prognosis, Korszun et al suggest that SSRIs may not alter the mechanisms through which depression leads to increased cardiovascular morbidity and mortality. However, two other explanations may be more plausible. First, the effects of antidepressant treatment on depression itself are not strong enough. In both the ENRICHD and SADHART trials, the response rates of patients in the active treatment arm hardly exceeded those of patients receiving usual care or placebo. Second, the cardiotoxic effects of depression are limited to patients for whom antidepressant treatment is not effective (Grace et al, 2005; de Jonge et al, 2006a). We have shown that the cardiotoxic effects of depression are concentrated in incident post-myocardial infarction depression, whereas results from the SADHART study have indicated that sertraline is only effective in non-incident depression (of interest, Korszun et al mention mechanisms related to recurrent depression, which appears not to be cardiotoxic). If antidepressant treatment is only effective in non-cardiotoxic depression, no effects on cardiovascular prognosis can be expected.

Shetty raises ethical concerns about our study, because we used Zelen’s method of randomisation. Controls were not told about their diagnosis of depression and, as argued by Shetty, therefore may have ‘missed’ an offer of adequate treatment. However, we feel that in 1999, when the study started, Zelen’s method was both scientifically useful and ethical because no controlled comparative studies had yet investigated the clinical efficacy and safety of antidepressant drugs in depression post-myocardial infarction. At that time, the proportion of myocardial infarction patients with depression who were treated for their post-myocardial infarction depression was negligible. In addition, serious concerns existed about the safety of antidepressant drugs in cardiac patients. Moreover, in our study patients with a significant risk of suicide or severe depression were excluded from randomisation and referred for psychiatric treatment outside the study. Finally, all patients received usual care, i.e. had cost-free access to all usual treatment facilities such as normal cardiac rehabilitation programmes and healthcare from family physicians. We therefore feel it was ethical to use Zelen’s method in our study and scientifically useful as our control patients were truly representative of patients receiving usual care.

We agree with Dr Kho that we need to develop new treatments for depression post-myocardial infarction, but believe it is premature to consider electroconvulsive therapy (ECT) as an effective alternative. In our experience those types of depression that are least similar to those seen in general psychiatry (i.e. incident depression occurring for the first time after myocardial infarction; de Jonge et al, 2006b) and those that are dominated by feelings of exhaustion rather than negative self-esteem or suicidality (de Jonge et al, 2006a) are the most cardiotoxic. To our knowledge the mechanism(s) explaining this remain unclear. Similarly, it is not known whether ECT is effective in these subtypes (although it appears that antidepressive medication is not). In fact, the studies mentioned by Dr Kho suggest increased rather than decreased cardiovascular events.

New, effective treatments for depression post-myocardial infarction will improve quality of life but perhaps also survival, as rightfully argued by Dr Kho. Carney et al (2004) showed that responders to antidepressive medication had a better cardiovascular prognosis than non-responders,
Substance misuse disguised as ADHD?

Attention-deficit hyperactivity disorder (ADHD) is a rather novel disease in adults. It has drawn increasing attention and at present there is no deficit of studies of ADHD in adults (Fayyad et al, 2007). Several studies have shown a considerable risk of co-occurring substance misuse in adults given the diagnosis of ADHD (Anonensen, 1999; Wilson, 2007). Symptoms of ADHD seem to hamper success in methadone maintenance treatment (Kolpe & Carlson, 2007). Fayyad et al indicate in Table 5 that in 99% of cases adult ADHD occurs first in patients with a co-occurring substance use disorder but this is not commented upon in the discussion part of their paper. Respondents were classified retrospectively as having met full ADHD criteria in childhood. To ascertain the presence of ADHD in adulthood respondents were asked a single question only, whether they continued to have problems with attention or hyperactivity.

In Norway we have an impression that people with substance misuse tend to ask for a diagnosis of ADHD, as this may lead to better treatment within the psychiatric care system. The finding of Fayyad et al of higher prevalences in high-income countries, with purportedly better services for the treatment of ADHD, may be an indication of common presenting symptoms in substance use disorder and ADHD. Could the authors have observed symptoms and behaviour related to substance misuse and not ADHD?

Authors’ reply

Dr Berg raises the possibility that respondents in our surveys who reported persistence of ADHD in adulthood might actually have had symptoms caused by some other disorders, such as alcoholism, that are more stigmatising and less likely to be treated as ADHD. Such respondents might consciously have provided incorrect information in an effort to avoid stigma and to increase their chances of receiving treatment. Dr Berg states that such machinations occur in his country. This is an important point in view of the stigma associated with mental disorders and the fact that some healthcare systems discriminate against certain diagnoses. Mental health professionals need to increase their efforts to raise awareness and address these problems.

That said, it strikes us as implausible that our findings are importantly affected by the sort of bias proposed by Dr Berg. First, the World Mental Health surveys are community epidemiological surveys in which no treatment is provided. Second, in a number of the participating countries ADHD is not commonly recognised as an illness, making it unlikely that community respondents would have the sophistication to seek out this diagnosis. Third, we carried out in-depth clinical reappraisal interviews with a probability sub-sample of respondents who reported adult persistence of ADHD. We excluded respondents if concerns existed that another diagnosis might be primary. Although it is possible that some respondents were so familiar with ADHD that they tricked our experienced clinical interviewers, we consider it unlikely that this was widespread. Fourth, treatment-seeking was low in most World Mental Health surveys. When it occurred, the reason for seeking treatment was not ADHD but a comorbid disorder.

Irrespective of whether the type of bias Dr Berg suggested exists in epidemiological surveys, our results imply that clinicians should look more seriously for ADHD in their adult patients than they have before. As more physicians screen for ADHD among adults presenting for treatment of other psychiatric disorders, the extent to which untreated adult ADHD exists among help-seekers will become apparent.


R. de Jonge Department of Internal Medicine and Department of Psychiatry, University Medical Center Groningen, The Netherlands. Email: p.d.jonge@med.umcg.nl

J. P. van Melle Department of Cardiology, Thoraxcenter, University Medical Center Groningen, The Netherlands

J. Ormel Department of Social Psychiatry and Psychiatric Epidemiology, University Medical Center Groningen, The Netherlands

doi: 10.1192/bjp.191.5.456a

R. C. Kessler Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA. Email: kessler@hcp.med.harvard.edu

doi: 10.1192/bjp.191.5.457
Heroin-assisted treatment: no difference in treatment retention

Haasen et al (2007) report highly significant findings from their trial of heroin plus methadone maintenance. A small problem is that the heroin plus methadone group were, to a large extent, self-selected, with only 2.3% failing to initiate treatment in this group. The methadone group, on the other hand, had a retention rate of 56.3%, log rank test statistic of 14.1, P < 0.001. Therefore, it is incorrect to say that ‘retention rates were possibly different insignificantly’: the difference is certainly less, but still significant.

Factors in those who repeatedly self-harm

We read with interest the article on young people who self-harm (Young et al, 2007) and feel that the outcome of factors considered would have been more viable if a further subgroup analysis was performed in those patients who repeatedly self-harm. A significant amount of our time is taken up by people who self-harm repeatedly. This subset of clients are often entrenched in their behaviour patterns and use services disproportionately. Existing studies have not adequately analysed factors responsible for repetition of self-harm and we feel that Young et al missed an excellent opportunity to investigate this, albeit in a younger age group.

An analysis of our data from the Integrated Care Pathway (Rajwai & Gash, 2006) showed repetition rates of 40% for 2004, 42% for 2005 and 43% for 2006 of all our referrals each year. This means that 18% of our patients in 2004, 18.9% in 2005 and 19.2% in 2006 were responsible for the above statistics year on year. These data are from adults of working age and only include repetition in the same calendar year. About 13% of our referrals are under 21, and 18% of those are for repetitions of self-harm. Hence a small proportion of our clients are responsible for a large proportion of our work.

Our data support Young et al on the lack of a gender bias in the prevalence of self-harm. Females comprised 50.2% of our referrals in 2006 but only 49.0% of those repeating self-harm. The old myth of a higher proportion of females self-harming was not borne out by our statistics, although we considered the entire adult age group.

We would be interested to know whether the results of Young et al would be different in the subgroup with repeated self-harm.
Kripalani we agreed a three-way classification of repeated self-harm among young people: repeated self-harm (19 out of 89), with self-harm both in the past and currently or using several (three or more) methods (since it is unlikely that multiple methods of self-injury refer to a single incident); a single incident (17 out of 89), with an explicit statement of a transient incident; unsure (53 out of 89), which constituted the remainder. The crude repetition rate of 20% is typical for self-harm (Bennewith et al., 2002). We proceeded to re-analyse the data from our original paper for repeat self-harm (results available on request).

In summary, we can confirm that repeated self-harm was unrelated to gender, or social class of origin, but was related to current labour market position, with youth outside the labour market more likely to self-harm repeatedly. Young people who repeatedly self-harmed were more likely to use all methods except taking pills and more violent methods, which were common to all groups. Those who repeatedly self-harmed were far more likely to do so to relieve negative emotions (anger, anxiety or to punish themselves), but self-harm with intention of killing oneself was common to all groups. Taken together this confirms that those who repeatedly self-harm are more likely to use self-injury as a coping mechanism. With regard to service use, those young people were nearly twice as likely to have used emergency services and over three times as likely to have used psychological services from the age of 11.

This suggests that young people and adults who repeatedly self-harm are heavy users of both health services in general and psychiatric health services in particular, and is compatible with the assertion of Kripalani et al. that a small proportion of clients may account for a large proportion of resources. Distinguishing between repeated and other forms of self-harm could provide useful clinical information, provided that both researchers and clinicians can agree on a clear definition.

**Attachment disorders: an evolutionary perspective**

In a large twin study Minnis et al. (2007) have demonstrated that attachment disorder behaviours can be differentiated from other common childhood emotional and behavioural disorders and appear to be strongly genetically influenced, particularly in boys. The authors also point out that, even in a population of children that was probably healthier than the general population, behaviours suggestive of attachment disorder were identified. Conventional aetiological factors are addressed but the paper would have benefited from the inclusion of an evolutionary perspective. Evolutionary or Darwinian psychiatry examines, among other things, the potential for adaptive benefits to pre-programmed psychobiological mechanisms (e.g. depressive symptoms or attachment disorders) that are sometimes incorrectly viewed as being simply abnormal or pathological (Abed, 2000).

It was surprising that Minnis et al. made no reference to Bowlby's seminal work (Bowlby, 1958) in the area of attachment. Bowlby's perspective on attachment was an evolutionary one, in that he viewed the associated behaviours as representing evolved and adaptive psychobiological mechanisms, protecting the child from predators and the many other dangers prevalent in our ancestral environment. This 'adaptationist' perspective could have been explored by Minnis et al. when considering why attachment disorder behaviours occurred at all in this healthy non-clinical sample.

Chisholm (1996) and Belsky (1997) proposed in more recent years an integration of life history theory (Levins, 1968) and attachment theory. Chisholm (1996) argued that, in life history theory, life cycles constitute evolved adaptive strategies. Furthermore, individuals must prioritise the allocation of their time and resources to different components of reproductive fitness (e.g. growth, mating or parenting). Therefore, the sexual strategy employed by parents (e.g. low investment in large numbers of offspring or vice versa) is an integral component of the child's early environment. Belsky (1997) argued that secure attachment in children functioned to promote a strategy of high-investment parenting, and avoidant attachment (child showing indifference to parent) as representing an adaptation to parental unwillingness to invest (e.g. when the parent invests instead in a short-term mating strategy with relatively little investment in individual offspring).

The anxious/ambivalent style of attachment evolved in response to parental inability (e.g. through illness) to invest, and fostered a 'helpers at the nest style' in the children, whereby children would cooperate in rearing siblings. For example, Turke (1988) demonstrated (independent of attachment disorders) that women from the Micronesian atoll of Ifaluk were likely to have significantly larger families when their first-born was female: an anxious/ambivalent attachment style may further accentuate such behaviour in female children, perhaps explaining in part the gender differences in attachment disorders raised by Minnis et al.

These are merely a few examples of the insights that evolutionary psychiatry can provide. In the total absence of such an evolutionary perspective, one is reminded of Abed's (2000) cautionary comments: 'In recent years psychiatry has attempted to circumvent such problems by engaging in an atheoretical research enterprise involving gathering masses of data and calculating sophisticated statistical associations. However, such an endeavour of itself cannot generate a scientific discipline, for science is a method of discovering the world and not simply a body of facts'.


H. P. O'Connell Clare Mental Health Services, Lisdoonvarna, County Clare, Ireland. Email: hpoconnell@yahoo.ie doi: 10.1192/bjp.191.5.459
We were interested to read Dr O’Connell’s response to our paper. Unfortunately we did not have space to do justice to a discussion of reactive attachment disorder from an evolutionary point of view, although we agree that this is an important theoretical perspective. Dr Minnis first became interested in reactive attachment disorder when working as an orphanage doctor in Guatemala. Most of the children there displayed symptoms of the disinhibited form of the disorder and it seemed clear that these behaviours were adaptive in a setting where primary attachment figures were lacking. We have touched on the maintenance of these behaviours from an evolutionary perspective in a previous paper (Minnis et al, 2006).

Dr O’Connell also points out that we did not engage in a discussion of attachment theory, or the work of John Bowlby (Bowlby, 1973). We do not wish to underestimate the crucial role of Bowlby’s work in advancing our understanding of childhood development, however, we were unable to do justice to the complex interplay between attachment patterns and reactive attachment disorder within the space allowed. This important topic is the focus of our previous publication (Minnis et al, 2006). In short, children can be securely attached while suffering from reactive attachment disorder and children suffering from the disorder have difficulties in various domains of early development, not simply the domain of attachment (Richters & Volkmar, 1994; Green & Goldwyn, 2002). Research into reactive attachment disorder is in its infancy and is a field ripe for exploration on a number of fronts. 


H. Minnis Section of Psychological Medicine, University of Glasgow, Glasgow, UK. Email: h.minnis@clinmed.gla.ac.uk

J. Reekie, D. Young, A. Gray Department of Statistics and Modelling Science, University of Strathclyde, Glasgow

T. O’Connor Laboratory for the Prevention of Mental Disorders, University of Rochester, New York, USA

A. Ronald, R. Plomin Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King’s College London, UK
doi: 10.1192/bjp.191.5.460

One hundred years ago

Stereotypy in Dementia Praecox
[Étude clinique sur la stéréotypie des démences précoces]. (Arch. de Neurol., March, 1905.) Dromard, G.

Stereotyped movements are not limited to dementia præcox, but occur in secondary dementias, and also in some systematised delusional states. Many followers of Kraepelin, however, would bring these last largely under the head of dementia præcox.

1. Classification:

(A) Akinetic stereotypy. – Attitudes, either of the whole body or of an individual member. Refusal of food, and mutism – though usually referred to negativism – are sometimes examples of akinetic stereotypy – e.g., a patient who refuses to eat, but who offers no resistance when fed with the nasal tube.

(B) Kinetic stereotypy. – Movements, which may be further subdivided into stereotypes of (1) speech – e.g., neologisms, constantly recurring words and phrases, modes of intonation; (2) writing; (3) expression – e.g., grimaces; (4) walking; (5) complex stereotypies – e.g., special modes of sitting, eating, or dressing. Masturbation is sometimes referred to in this group.

2. Evolution. – The stereotypies of the acute stages of the disease must be distinguished from those of the terminal period.

(A) Primitive stereotypies. – The prolonged attitudes and repeated movements of katatonia. The failure of physiological plasticity, the impeded psychical process of the katatonic, are probably of toxic origin. The movements are angular, jerky, awkward, like those of a mechanical toy. This type of movement tends to disappear in later stages.

(B) Secondary stereotypies. – Those of the terminal period. These are not to be correlated with a functional failure in the cells, but with an organic defect, the result of the previous toxin. These movements therefore arise from imperfect intercellular connections – that is to say, a state of disaggregation. Secondary stereotypies are the residue of acts which, though once adapted, conscious, and voluntary, are now purely automatic. The original idea is often to be found in the hallucinations and delusions, accompanied by profound emotional colouring, which occur in the early period of the disease. “Professional” acts also frequently form the basis of subsequent stereotypies, but some automatic movements must be regarded as of fortuitous origin.

Although the secondary forms imply a more advanced stage of disease than the primary, they nevertheless may occur comparatively early, often contemporaneously with the latter. This is analogous to the co-existence in a tissue of inflammation and sclerosis.

Secondary stereotypies tend to become reduced in number as time goes on, those remaining being usually those first formed.

3. Clinical value. To preserve the value of stereotypy as a clinical sign, the meaning of the word must be strictly limited. A repeated action is not stereotypy if it is still joined to an idea. Acts committed under the influence of obsessions, the conjurations of paranoia, etc., must therefore be excluded.
Stereotypy is far commoner in dementia praecox than in other forms of mental disease. Primitive stereotypies are more frequent in katatonia, secondary in hebephrenia and dementia paranoides. Stereotypy serves to distinguish the excitement of dementia praecox from that of manic-depressive insanity and general paralysis. It occurs early in dementia praecox, late in systematised delusional insanity. It is also of service in distinguishing the terminal stages of dementia praecox from other terminal dementias.

As regards prognostic value, secondary stereotypies are of grave import; in primary stereotypy the outlook is less gloomy, especially if other signs of active katatonia are present. Nevertheless, in the so-called “cured” cases of dementia praecox a tendency to stereotypy persists. This tendency may be utilised in teaching the patients simple machine-like occupations.

**REFERENCE**

Journal of Mental Science, January 1907, 183.

Researched by Henry Rollin, Emeritus Consultant Psychiatrist, Horton Hospital, Epsom, Surrey
doi: 10.1192/bjp.191.5.460a
Book reviews

EDITED BY SIDNEY CROWN, FEMI OYEBODE and ROSALIND RAMSAY

Fish’s Clinical Psychopathology: Signs and Symptoms in Psychiatry (3rd edn)


The above can be variously reviewed. At its blandest, the reviewer could say that one should be grateful to the Royal College of Psychiatrists for the idea of putting out this third ‘edition’. In the same vein, it could be added that the book needed updating and that the editors have done a good job; and then the crowning platitude included: it should be on the bookshelf of all UK trainees. This waffling, however, would be unfair to Frank Fish, Max Hamilton and indeed to current trainees, all of whom deserve better than that.

A new edition should have taken the opportunity to explain: (a) the meaning, history and significance of this work; (b) what clinical psychopathology is and what its role should be in current psychiatric research; and, most of all, (c) it should have included an essential excursus on what it means to ‘update’ a book on clinical psychopathology. I well remember Max Hamilton (I was then his lecturer at Leeds) worrying about how this could be done: would that entail changing the ‘descriptions’ that Frank had ‘got wrong’? ‘Adding’ symptoms missed out or recently ‘discovered’? These questions are as valid today as they were in 1974.

Hamilton was careful to keep (almost intact) the introduction because he considered it as one of the most important parts of the book. Where is it now? Equally respectful was he of the chapter on classification as it dealt with psychiatric taxonomy and not with the latest classification in the market. In the third edition, this chapter has been distorted by replacing the classical Störring & Schneider conceptual taxonomy with a classification taken from an American manual. The same can be said of the bibliography to which Hamilton added but did not subtract: the third edition has replaced all the classical references by ephemera thereby leaving many of Fish’s claims un referenced.

To decide on the appropriateness of the current changes, a serious review (which this book needs) should collate all editions. Just one example of an idle addition should suffice. Fish and Hamilton say all that can be said about the clinical psychopathology of hallucinations. The current edition adds: ‘SCAN (WHO, 1998) defines hallucinations as “false perceptions”. ’. In what way does this gem improve things? Has that ‘definition’ not been around since 1817? And so on and so forth.

The publication policy of Gaskell remains a mystery to most of us but if this work is typical of it then we should be concerned. What was wrong with reprinting the original or Hamilton 1974 effort? Ideally, of course, the book should have been contextualised particularly in relation to the blending of the newly arrived German ideas and the remnants of the in-house psychodynamic approach, as epitomised in the books by Hart and Nicole, that had dominated British psychopathology until the Second World War.

My advice to anyone curious about this book is: borrow the 1967 edition! It will contain strange concepts and names but this is OK as that might induce you to read up further. If thus, the old book will have achieved its objective.

German E. Berrios University of Cambridge, Department of Psychiatry, Addenbrooke’s Hospital, Box 189, Hills Road, Cambridge CB2 2QX, UK. Email: gebl1@cam.ac.uk
doi: 10.1192/bjp.191.5.462

Most psychiatrists of a certain age possess a luridly pink slim volume known to them simply as ‘Fish’. Frank Fish’s first edition is now 40 years old; the second edition, prepared by Max Hamilton, first appeared in 1974 and was last printed in 1985. Does a revised third edition have anything to offer a new generation of psychiatrists?

It has certainly managed to keep some of the main strengths of the original. The vivid clinical descriptions capture something of the strangeness that abnormal thoughts and experiences must have for those who suffer them. This should be helpful to exam-weary MRCPsych candidates, and their supervisors, in demonstrating that the systematic assessment of mental symptoms is both fascinating and rewarding. The chapters on disorders of emotion, disorders of the experience of self and (unsurprisingly in view of the senior author’s interests) on personality disorders are up-to-date, well referenced and provide lucid summaries both of new evidence and of areas of persisting controversy (such as the status of borderline personality disorder).

Some of the other chapters have not been updated as extensively: all but one of the references in the chapter on classification, for example, are from before 2000. This chapter would also have benefited from more critical discussion of the currently used classificatory systems and the challenges for DSM–V and ICD–11. Although symptoms are lucidly described throughout the book, there is little guidance on how to elicit them and the cultural dimension is all but ignored. This is particularly striking in the appendix on ‘psychiatric syndromes’ which has a single paragraph on ‘culture-bound disorders’. Post-traumatic stress disorder (PTSD) is omitted entirely, which is particularly surprising given the vivid and varied range of psychopathology with which PTSD victims can present. The appendix on ‘defences and distortions’ provides clear, exam-friendly definitions but fails to place them in the
context of the psychodynamic and cognitive frameworks in which they belong.

Overall, the authors have made a brave, if doomed, attempt at achieving the irreconcilable aims of preserving the character of a book which is the product of its (now quite distant) time while also writing something of practical use for today’s psychiatrists.

Cornelius Katona Kent Institute of Medicine and Health Sciences, University of Kent, Canterbury CT2 7PD, UK. Email: c.katona@kent.ac.uk
doi: 10.1192/bjp.191.5.462a

Severe Dementia
Edited by Alistair Burns & Bengt Winblad. John Wiley & Sons. 2006. 269pp. £75.00 (hb).
ISBN 0470010541

In their introduction, the editors point out that much research and clinical attention is being directed at early diagnosis and treatment of the mild and moderate stages of the disease and, as a consequence, severe dementia is a relatively neglected area. A motivating factor behind this book was to redress this imbalance and bring together the key issues and current knowledge of severe dementia. The contributions of experts from a variety of backgrounds have succeeded in this.

The early chapters of the book cover assessment, diagnosis, brain chemistry and molecular pathology. These chapters are not entirely specific to severe dementia, but do give an up-to-date account of current knowledge. Similarly, the section on clinical features includes material relevant to the earlier stages of the illness, but the chapters on staging and function in severe dementia include considerable detail and are rich with information.

The final section covers the management of severe dementia and includes chapters on drug treatments, non-pharmacological interventions, palliative care and health economics. Those on drug treatments highlight the relative dearth of robust research in severe dementia, but the topics are comprehensively covered. Drug treatments for behavioural and psychological symptoms are well covered and are followed by a general overview of non-pharmacological treatments and then an interesting chapter detailing the behavioural and environmental interventions of the Seattle protocols. There follows a short chapter on ‘Care by families’ – research in this area is relatively scant but what there is, and the issues brought to the fore by the authors, are very pertinent. The remainder of the book is very much specific to severe dementia, with thought-provoking chapters on person-centred care, end-of-life issues and narrative ethics. The final chapter on health economics reminds the reader that the severe stages of dementia are the most costly, yet little is known about resource use and cost-effectiveness of interventions for the late-stage illness.

In summary, this is an excellent book that truly brings some focus back onto the nature of, and issues associated with, severe dementia. It will be a valuable resource for specialist clinicians and those directly providing care to people with severe dementia, such as general practitioners and staff of nursing homes.

Peter Bowie Sheffield Care Trust, Fulwood House, Old Fulwood Road, Sheffield S10 3TH, UK.
Email: peter.bowie@blueyonder.co.uk
doi: 10.1192/bjp.191.5.463

Binge Britain: Alcohol and the National Response

As the title announces, this book concerns itself largely with the pattern of drinking in Britain, past and present, and seeks to comment on the government response to problematic alcohol use in terms of policy-making, legislation and its enforcement. Its publication is timely, in so far as it follows the 2004 National Alcohol Harm Reduction Strategy for England, which outlines interventions to prevent, minimise and manage alcohol-related harm. In 2005 the Alcohol Needs Assessment Research Project found that 23% of the population aged 16 to 64 drink hazardously or harmfully (7.1 million in England) and a further 1.1 million are dependent. Furthermore, 21% of men and 9% of women are binge drinkers. Alcohol problems are estimated to cost the taxpayer more than £20 billion per annum, and alcohol is implicated in 30,000 hospital admissions, 70% of accident and emergency attendances and 22,000 premature deaths.

Binge Britain is certainly a readable book. It provides an informative historical overview that examines factors influencing alcohol use and the social consequences of alcohol consumption. It explores the role of public attitude and influence of the alcohol industry in contributing to and maintaining the current binge style of drinking in Britain. The book also highlights the growing concerns regarding the escalating use of alcohol among British women.

Overall, I found it a useful source of historical and social information. I was disappointed that the potential physical and psychological consequences of excess alcohol consumption are not explored in more detail. I also felt that with the authors’
keenness to impress the flaws of the government’s illogical choices following their National Alcohol Harm Reduction Strategy (such as the significant evidence against the extension of licensing hours), an objective discussion seemed to have been sacrificed. It is, perhaps, a little hopeful to expect a group of scientific experts, even with a strong evidence base, to outweigh an industry worth billions to the government.

All in all, if you are looking for a summary of Britain’s obsession with alcohol over the ages, including more recent trends and political policies, *Binge Britain* is a worthwhile read.

Elizabeth Furlong Mary Street Community Drug Team, Balsall Heath, Birmingham B12 9RN, UK. Email: elongfur@yahoo.co.uk doi:10.1192/bjp.191.5.463a

Dynamic Security: The Democratic Therapeutic Community in Prison


If, as various sages have asserted, the state of a nation’s prison system is a reflection of that society’s socio-political health, where does that leave the UK? Well, frankly, in the doldrums, in need of little short of a revolution. Current penal policy persists in a largely hostile and non-rehabilitative attitude to the offender, who is invariably warehoused in one of the increasing number of our increasingly overcrowded, prisons. But, as in the larger matter, there are (quantiav)ly small beacon of light within this rotten system, which seek to understand and address the many factors which play their part in the personal and societal failures and tragedies which result in criminality. The subject matter of this volume is such an illumination: specifically, the combination of activities needed to run a democratic therapeutic community within the prison system.

This well-edited book is written by psychologists, prison governors, therapists, psychiatrists and researchers. It contains different sections: three chapters on theories of criminality; three on the history of the therapeutic community in prisons – especially well done by Newell & Healey – which trace the pervasive influence of Maxwell Jones and the Henderson Hospital experience, and three in a section covering methods and practice. In one of these, Alan Miller reviews recent initiatives for resettlement and support of prisoners after prison discharge. It is a sobering and depressing fact that 95% of inmates move on from prison therapeutic communities back to the general prison, with all its pressures towards reaffirmation of the criminal identity. So, post-therapy after-care needs to include initial support in surviving (again) the depredations of prison, and only thereafter, support in the outside community.

In my view this return to prison – quite unnecessary other than for bureaucratic reasons – is but one example of the common phenomenon of sabotage of good work undertaken in the prison. The internal saboteur of creative work by offender and staff is a crucial dynamic in understanding the stagnation of the individual and the institution, and to help overcome it. Neither this, nor any psychodynamic theory, receives attention by Day in an otherwise competent overview chapter on psychological theories of criminality; nor is it addressed in another vivid section on psychodynamic aspects.

There is a helpful section on managing the therapeutic community – a difficult task in a frequently uninterested and sometimes antagonistic institutional structure. The old canard that the application of the therapeutic community model of treatment, of its essence democratic and free, is therefore unsuitable to the coercive situation of the prison, is ably dealt with by several authors in this section. There follows a chapter on audit and accreditation, now required for all treatment programmes within prisons, and an account of a body created jointly by the Association of Therapeutic Communities and the Royal College of Psychiatrists Research unit, called the ‘Community of Communities’ – a voluntary network of peer review and quality assurance for therapeutic communities of whatever hue, in whatever setting, in the UK or abroad.

The penultimate chapter contains four rather brief accounts by individuals who have experienced therapy within a prison therapeutic community. They are, clearly, selected and merely illustrative but they make their points well, and are not uncritical. One contains the comment ‘I think the massive overreaction in Grendon by Security in the last few years has damaged therapy’. This is a constant danger, common to all mental health provision and not just the forensic.

Why, then, does our dominant penal policy continue to be reparatively bankrupt as well as economically and humanly expensive, and further, why does our society and electorate continue to tolerate it? Almost the last words in the book from a Grendon prisoner may give us a clue:

‘We all do it, we all keep up a hard man front, we have to because if we don’t we’ll get crushed. We don’t want to, though, not always. There’s hundreds of us out there (in the prison system) who are dying to find some peace and security for once in their lives but we’re never going to be the first to say so, it’s too dangerous . . .’

This is familiar to those who work in the system. The contributors to this excellent volume know it well, and in different ways express their versions of trying to change this mind set. It is surely up to us to help change the reciprocal ‘macho’ attitude of politicians and the penal system itself.

Certainly, the paranoid attitude to offenders and their demonisation has the comfort of simplicity and retributive appeal. It may make us feel better but it is the cause of a continuing sink system. More disturbingly, perhaps, for our own psychological purposes, we need to have this already marginalised group (over half of all prisoners are graduates of our equally awful care system) to further punish, neglect and vilify. All the more praise for the contributors to this volume.

Christopher Cordess University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK. Email: c.c.cordess@sheffield.ac.uk doi:10.1192/bjp.191.5.464
Depression and Physical Illness
Edited by Andrew Steptoe.
Cambridge University Press. 2007. 434pp. £45.00 (pb). ISBN 9780521603609

An important relationship between depression and medical disorders has long been recognised and variously conceptualised according to prevailing beliefs – from black bile in ancient Greece to psychoanalytic theories in which the body is a theatre of the mind and a myocardial infarction can result from unexpressed emotions. For most of the 20th century the term psychosomatic medicine encompassed an interaction between the psychological and the physiological. However, in many ways this term divided between body and mind even wider, with many patients and clinicians continuing to regard a psychological condition as not real. Depression is still described as psychological rather than biological, yet, in the light of our current knowledge, what could be more biological than a neuroendocrine brain disorder with multiple manifestations in various organ systems?

In this excellent new book, Andrew Steptoe brings together a wide group of experts to give us a 21st-century view of the associations between depression and physical illness. While addressing psychological aspects of depression, it also presents a carefully balanced view that critically evaluates the biological underpinnings of the disease associations. (Interestingly, the term psychosomatic does not even appear in the index.)

The book is divided into three parts. The first gives a clear account of the importance of the definition of depression and its methods of measurement. It also provides a comprehensive overview of how psychosocial factors, such as low socio-economic status and education, predict not only depression but also affect physical risk profiles. A range of specific health problems is covered in the second part and in the third possible underlying biological and behavioural processes are explored.

The most robust links are between depression and coronary heart disease and this is reflected by the inclusion of three chapters giving a balanced and thorough presentation of the evidence that individuals with depression are more vulnerable to heart disease and even if their depression can be successfully treated this will not necessarily improve the course of the cardiac disease. The evidence linking depression with other medical disorders, such as diabetes and cancer, is not as strong but is nevertheless consistent. The chapter on pain and depression gives a fascinating overview of these two conditions as related symptom complexes associated with neuroendocrine and immune activation. Similar links are described for other disorders and these findings are integrated in the final section in working models that indicate future opportunities and possible pitfalls in this field of study.

I recommend this book highly to all mental healthcare professionals and my only criticism is that I would have liked to read more about potential treatment approaches but, as this field continues its rapid expansion, we can look forward to a larger, later edition. But most of all, I would recommend this book to other medical specialties – on checking the contents of the latest editions of several prominent textbooks of medicine, I could find no reference to the role of depression despite the overwhelming evidence presented in Depression and Physical Illness.

Ania Korszun  Department of Psychological Medicine, University of Wales, Health Park, Cardiff CF4 4XN, UK. Email: korszun@cardiff.ac.uk doi: 10.1192/bjp.191.5.465

Formulation in Psychology and Psychotherapy: Making Sense of People’s Problems

Most clinical psychologists and psychotherapists respect case formulation as an aid to good practice. For many psychiatrists, it remains a source of anxiety and confusion. Although the former are this book’s natural audience, I think it has much to offer inquiring psychiatric trainees. Comparative accounts of the psychotherapies can provide brief portraits that don’t convey what their relative strengths and weaknesses are. Using the vehicle of the case formulation, this book often succeeds in describing and demonstrating key differences in how clinicians using different models think. In covering a variety of perspectives – not only cognitive–behavioural, psychodynamic, systemic and integrative, but also social inequalities and social constructivist viewpoints – each psychologist contributor has been asked to produce specimen formulations for two case vignettes: a young man expressing paranoid fears and an anxious 9-year-old girl with developmental problems (although some pass on the latter).

Its success is uneven, however. Some chapters, such as those on cognitive–behavioural therapy and systemic family work, are exemplary introductions to formulation within these models. Other authors are diverted into spending unnecessary words on outlining the principles of their model at the expense of its approach to formulation. Often, little attention is paid to how a formulation would be used to facilitate treatment within a particular model, in favour of its purely descriptive functions. The book also makes surprisingly few references to the considerable research literature on formulation. Several well-known, research-based systems are ignored altogether, as are two major international attempts to systematise psycho-dynamical formulation. The book’s tone may also deter some readers. As far as I could tell, amid repeated references to ‘problems’ and ‘distress’ as the basis of client’s suffering, the word ‘illness’ fails to appear. The omission can arouse suspicion as to how fully the impact of pain, dependence and loss, as well as stigma, is appreciated. The editors’ credulous stance in relation to the diatribes of Jeffrey Masson may also undermine some readers’ confidence. However, the book’s occasional infelicities are offset by consideration of areas of a patient’s positive strength and resilience within formulation, as well as a healthy wariness concerning the dangers of allowing formulation to unduly restrict the ability to see what is in front of us.

Chris Mace  St Michael’s Hospital, Warwick
CV34 3QW UK. Email: c.mace@warwick.ac.uk doi: 10.1192/bjp.191.5.465a
This book highlights the renewed recognition of the value of spiritual dimensions of health with the growth of ethics. Science and religion, as grand narratives, are being replaced in the post-modern era by personal narratives. Although individualised care needs to be developed with a scientific evidence base, it should be offered in a personalised therapeutic relationship. A doctor needs to understand a patient’s problem not only from a scientific perspective but also from one of faith. It requires a personal encounter and not a computerisable communication. This should involve the whole person of the caregiver who also has access to informatics. It is like joining hands: the hand of scientific competence and the hand of personal communication. Therapeutic relationships can also induce true biological relationships can also induce true biological hand of scientific competence and the hand of the caregiver who also has access to communication. This should involve the whole encounter and not a computerisable com- nalised therapeutic relationship. A doctor needs to develop with a scientific evidence base, it should be offered in a perso nealised therapeutic relationship. A doctor needs to understand a patient’s problem not only from a scientific perspective but also from one of faith. It requires a personal encounter and not a computerisable communication. This should involve the whole person of the caregiver who also has access to informatics. It is like joining hands: the hand of scientific competence and the hand of personal communication. Therapeutic relationships can also induce true biological relationships can also induce true biological hand of scientific competence and the hand of the caregiver who also has access to communication. This should involve the whole encounter and not a computerisable com

Reconceiving Schizophrenia
Edited by Man Cheung Chung.
K.W.M. (Bill) Fulford, & George Graham.
Oxford University Press.
2007. 341pp. £29.95 (pb).
ISBN 9780198526131

As a medical student I recall being told by a geriatrician that the longer they practised medicine, the harder they found it to confidently diagnose Parkinson’s disease. At the time I was a little perplexed by this but now I begin to feel similarly about schizoph renovation. The confidence I had in schizophrenia having a clear-cut clinical presentation, mapping onto a similarly discrete and specific pathophysiology, evaporated in my first few weeks of psychiatric training. Having been fortunate to have worked predominantly both clinically and academically in psychosis, this scepticism has been further compounded. Reconceiving Schizophrenia is part of the successful International Perspectives in Philosophy and Psychiatry series. It contains 16 chapters, all on schizophrenia, utilising philosophy to examine our assumptions and ways of understanding this most emblematic disorder for psychiatry.

The chapters are not formally subdivided into themes: introductory and review chapters open the volume. Chung’s review is a helpful resource for any researcher interested in more philosophical approaches to schizophrenia and amazingly manages to distil the literature, from phenomenological psychiatry to psychiatric classification, in 34 pages. This is followed by four chapters exploring the role phenomenological psychiatry continues to play in understanding major mental illness. The latter half of the volume is more analytic and anglophone, with outstanding contributions from Hamilton and Stephens and Graham on delusions. Gillet offers a fascinating account of psychosis, drawing on Kant’s Anthropology, and struggles with how meaning in schizophrenia can both be private and yet, in some senses, communicable. Poland’s chapter is a timely discussion of ‘the schizophrenia concept’. It often seems that the idea of schizophrenia that

Medicine of the Person: Faith, Science and Values in Health Care Provision
ISBN 1843103974

This book is inspired by the work of Paul Tournier (1898–1986), a Christian doctor from Geneva, who was trained by a psychiatrist, Lechler. In Lechler’s daily meetings, when someone spoke, it was impossible to tell whether it was a doctor or a patient. The book has chapters on themes from faith traditions such as, Christianity, Judaism, Islam and Hinduism. Phrases like ‘touch wood’ or ‘cross your fingers’ are often used by health personnel and allude to the Cross, which may be disconcerting to patients from other faiths. Collectivism in Islam means that the basis of treatment should include uniting the person with the family. Ayurveda, a medical discipline developed in ancient India, incorporated the prevalent value systems. In Hindu thinking there is also the law of cause and effect and the goal of liberation from the cycle of rebirth. Contributors include eminent thinkers in their field and topics such as public health, neuroscience, general practice, home treatment and terminal care have all been included.

I feel the book is very timely and is likely to inspire further work with examples of good practice, particularly when medicine is being swamped with administration, technocracy, politics and management.

Kedar Dwivedi Northampton Child and Family Consultation Service, 8 Notre Dame Mews, Northampton NN1 2BG, UK. Email: kedar.dwivedi@ngh.nhs.uk
DOI: 10.1192/bjp.191.5.466

This book highlights the renewed recognition of the value of spiritual dimensions of health with the growth of ethics. Science and religion, as grand narratives, are being replaced in the post-modern era by personal narratives. Although individualised care needs to be developed with a scientific evidence base, it should be offered in a personalised therapeutic relationship. A doctor needs to understand a patient’s problem not only from a scientific perspective but also from one of faith. It requires a personal encounter and not a computerisable communication. This should involve the whole person of the caregiver who also has access to informatics. It is like joining hands: the hand of scientific competence and the hand of personal communication. Therapeutic relationships can also induce true biological relationships can also induce true biological hand of scientific competence and the hand of the caregiver who also has access to communication. This should involve the whole encounter and not a computerisable com...
anti-psychiatrists charge us with holding is one that is never held in practice. Indeed, it is perhaps a concept that one is disabused of as one learns the complexity of mental illness and the limits of science. Harré’s chapter is a fascinating meta-account of the discourses and grammars used when mental illness is discussed, and how the file-selves of psychiatric records come into being and the use they are put to.

The volume contains many thought-provoking and worthwhile contributions, with little overlap of content, and all of them deserve detailed consideration. It serves as an amazing achievement of conceptual rigor in thinking about schizophrenia.

Matthew Broome  Health Services Research Institute, Gibbet Hill Campus, University of Warwick, Coventry CV4 7AL, UK. Email: m.r.broome@warwick.ac.uk
doi: 10.1192/bjp.191.5.466a

Love and Loss: The Roots of Grief and its Complications

Colin Murray Parkes’ seminal book Bereavement: Studies of Grief in Adult Life, published in 1972, provided us with acutely observed accounts of women’s response to the untimely death of a spouse. The book became a classic. More than 30 years later Love and Loss provides further rich insights into the reactions of those who are bereaved. The ideas that Colin Murray Parkes knits together go beyond description to propose an explanation, rooted in attachment theory, for the nature of complicated responses to bereavement. The thinking expounded here is destined to become part of the accepted fabric of those working in this field and will undoubtedly prompt continuing debate and further research.

The volume takes the reader, step by step, on a journey that provokes us to consider the complex connections between childhood attachment patterns, parental nurturance, intimate relationships between adults and responses to bereavement. The combination of Parkes’ own research data, clinical case examples and ideas from the wider body of knowledge make for a multi-faceted and full-bodied text. The clinical examples bring the ideas to life, demonstrating lifespan and intergenerational influences, and make these transparent for the professional and non-professional reader alike.

The author does not flinch from venturing into sensitive areas. Not only is it tricky to research the grief of bereavement but this volume also threads its way into the labyrinths of love. The material is conveyed with characteristic compassion and reflexivity. The writing demonstrates tremendous respect for those whose early life experience distorts their ability to trust and leaves them struggling in the wake of loss.

The research study at the centre of the book has limitations in its sample and methodology. Standardised measures of attachment, grief and psychiatric symptomatology would have enhanced the validity of the results. Some of the novel messages of the book are based more on conviction drawn from thoughtful analysis and clinical experience rather than being fully supported by valid research data. However, this is the very delight of this volume. Parkes has an unequalled store of knowledge and experience. His thinking continues to develop and he has a wide overview of the territory. The book deserves to be read not for watertight evidence-based conclusions but for inspiration from the insights that can only come from open-minded analysis of extensive clinical experience.

Jan R. Oyebode  School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 8DL, UK. Email: j.r.oyebode@shsm.ac.uk
doi: 10.1192/bjp.191.5.467
From the Editor’s desk

PETER TYRER

DINING AT THE SLOW FOOD RESTAURANT

I have recently returned from a summer school in Denmark, where they not only enjoy their lifestyle but are more than happy to promote it. So we were reminded that Danes tend to start work ‘half an hour later than most other nations and, to compensate for this, finish half an hour earlier’. This gives them plenty of time for the delectable task of eating, and so, like the tortoise and the hare, slow food restaurants – indeed they are advertised as such – are catching up inexorably on the faster ones. In this issue I would like you to imagine you are eating at the slow food British Journal of Psychiatry restaurant and implore you to give a little more time to relish the articles and dwell on their messages.

In this context it is worth reminding our readers that one of the best known Danish specialties is smørrebrød, or the open sandwich. The Danes believe it is best to see exactly what you are eating and in the same way the Journal wants you to dwell on its pages and digest each precisely and slowly. You might start with support for the statement that the best tranquiliser in the world is a good square meal, as it is the Danes and the Swedes (please forget all that nonsense about every Swede being about as optimistic as Ingmar Bergman who have the least affective disturbance in old age (Castro-Costa et al, pp. 393–401). If you then move to the topical issue of cannabis and schizophrenia, we see each element of the smørrebrød exposed, with three genes separately examined for their role in the link. The answer appears to be ‘no’ for all three (Zammit et al, pp. 402–407), so spiking the claims that if you lack the ValMet polymorphism you can smoke cannabis indefinitely without risk of psychosis. Moving onto the next course from that flamboyant master chef Stefan Priebe and his European cooking team, you will see that, far from computers replacing good doctor–patient communication, they can enhance it (Priebe et al, pp. 420–426). Perhaps it is not surprising that focused patient-oriented interaction does so well, as other research we have recently published supports this (Johnson et al, 2003; Garety et al, 2006; Kendrick et al, 2006), but we still need to hammer home that our services should be focused on patient rather than manager satisfaction (Shipley et al, 2000) and that this offers an opportunity for outcomes to be used positively in promoting better care (Slade et al, 2006). In any case, computers are more logical than people, and clearly more suited to helping to treat schizophrenia (Owen et al, pp. 453–454).

Close and careful study of the exciting paper by Boisjoli et al (pp. 415–419) is also justified. Anything that brings disruptive boys back into a socially inclusive fold is worth promoting and the evidence that academic performance was promoted and criminal behaviour reduced over a 15-year period in their ‘preventive intervention’ is impressive. But what is the main component of this intervention’s three-pronged assault, or are all three needed? This brings us back to food again, or rather its avoidance, and ourselves a pat on the back for their efforts in successfully treating this condition and that in-patient services should be avoided. Is there a message here? Well, everyone knows how unattractive hospital food is to the average customer, and when this knowledge is combined with the expansion of Danish cuisine across the Oresund to Sweden, is it so surprising that the outcome of Swedish adolescents with anorexia nervosa has improved dramatically (Hjern et al, 2006)? So, please, be like the Danes, don’t put a piece of another journal’s indigestible verbal bread over your British Journal of Psychiatry smørrebrød, let it caress your taste buds and allow you to stay at the slow food restaurant till closing time.

SERVING UP THE FAST-TRACK OPTION

We introduced a fast-track option in the British Journal of Psychiatry a year ago. This, like many innovations in medical journalism, was introduced by The Lancet, who claim that for the really important papers they ‘can publish a peer-reviewed manuscript in the print journal in as little as 4 weeks of receipt, after full clinical/scientific and statistical review’. Nice aspiration, but I wonder how often this really happens. We will be reporting on our own fast-track results shortly, but in the meantime could I remind contributors that the following recent reasons given for the fast track route are not, shall we say, looked on with enthusiasm:

Try again, please.


