

## Bipolar disorder: clinical uncertainty, evidence-based medicine and large-scale randomised trials

JOHN GEDDES and GUY GOODWIN

**Background** The increasing use of the methods of evidence-based medicine to keep up-to-date with the research literature highlights the absence of high-quality evidence in many areas in psychiatry.

**Aims** To outline current uncertainties in the maintenance treatment of bipolar disorder and to describe some of the decisions involved in designing a large simple trial.

**Method** We describe some of the strategies of evidence-based medicine, and how they can be applied in practice, focusing specifically on the area of bipolar disorder.

**Results** One of the key clinical uncertainties in the treatment of bipolar disorder is the place of maintenance drug treatments and their relative efficacy. A large-scale study, the Bipolar Affective Disorder: Lithium Anticonvulsant Evaluation (BALANCE) trial, is proposed to compare the effectiveness of lithium, valproate and the combination of lithium and valproate.

**Conclusions** Providing reliable answers to key clinical questions in psychiatry will require new approaches to clinical trials. These will need to be far larger than previously appreciated and will therefore need to be collaborative ventures involving front-line clinicians.

**Declaration of interest** None.

Although there is an ever-increasing demand for standardisation and improved quality in psychiatric treatment, it is extremely difficult for busy psychiatrists to keep up-to-date with clinically important advances. About 5500 potentially clinically relevant scientific papers are published annually: this is equivalent to reading 15 papers per day, which is clearly unrealistic for most clinicians (Geddes *et al*, 1999). This leads to a gap between research and practice, which is likely to be reflected by inappropriate variations in clinical practice.

### EVIDENCE-BASED MEDICINE

Evidence-based medicine was developed as a set of strategies derived from developments in information technology and clinical epidemiology, to assist the clinician in keeping up-to-date with the best available evidence (Geddes & Harrison, 1997). The aim of evidence-based medicine is to provide a set of tools to assist clinicians in decision-making. One of the key developments underpinning the feasibility of evidence-based medicine has been the development of the methodology of reviewing. Systematic review of primary research studies can critically examine the evidence for both established and new interventions and should provide reliable information to inform treatment choices. The advantage of the reviews contained in the *Cochrane Database of Systematic Reviews* is that, on average, they are of a higher methodological standard than other reviews (Jadad *et al*, 1998) and that, being primarily electronic, they can be continually updated as new primary studies are completed or identified.

Systematic reviews can also identify areas of substantial clinical uncertainty in which there is a dearth of good-quality research (Geddes, 1999). In the UK and elsewhere, mechanisms are being developed to feed back this clinical uncertainty to inform the

clinical research agenda, to ensure that scarce research funds are targeted at the areas of highest priority (Stein & Milne, 1999).

To make sure that the results of systematic reviews are introduced effectively into everyday clinical practice, it seems likely that it will be important to consider defined interventions concentrating on the problems raised in management by relatively coherent patient groups. Clinical epidemiology, which is one of the main methods of evidence-based medicine, has been called 'a basic science for clinical medicine' and can be seen to strengthen and enhance traditional medical approaches to the care of patients (Sackett *et al*, 1991). The introduction of evidence-based medicine is likely to lead to a renewed interest in accurate diagnosis and the efficient delivery of effective medical care for individual patients with defined problems. This approach is in contrast to the application of policies of service delivery based on ill-defined interventions for vaguely specified patient groups that have been fashionable recently, and have been described as a disaster for psychiatry in the 1990s (Freeman, 1999).

### Clinical uncertainty in the treatment of bipolar disorder

The drugs that are currently described as mood stabilisers include lithium and a number of anticonvulsants. Lithium, alone of these drugs, appears to meet all the conceivable criteria for stabilising mood. These include:

- (a) antimanic and antidepressant action in acute illness in bipolar patients;
- (b) evidence that efficacy in one pole of the illness does not produce switching to the other pole;
- (c) evidence for reduced mood variation in euthymia;
- (d) evidence for the prevention of new episodes of illness in long-term treatment.

The anticonvulsants were introduced in part because of the perceived analogy between bipolar mood disorder and epilepsy. This appears to have been remarkably prescient. However, we need to know more about the mechanism of action of these drugs before we can be confident that a particular pharmacological profile will predict efficacy. Despite merging research findings (Bowden *et al*, 2000), none of the anticonvulsants have yet been

established as securely as lithium was and pharmacology offers us little certainty in choosing between different treatment options. However, their cellular actions may offer important clues to pathophysiology, as described by Manji *et al* (2001, this supplement).

There are extreme variations in the prescription pattern of mood stabilisers in different countries. The USA is probably the most unusual. Several studies have documented the widespread and dramatic increase in the use of valproate (both alone and in combination with lithium), and the decrease in the use of carbamazepine and lithium monotherapy for the acute and maintenance treatment of bipolar disorder that has occurred in the USA since the early 1990s (Fenn *et al*, 1996; Citrome *et al*, 1998; Sanderson, 1998). For example, in New York State 15.5% of 18 668 psychiatric in-patients received valproate in 1994, whereas in 1996 valproate was prescribed to 34.1% of 12 444 patients (rate difference, 18.6%; 95% CI, 17.7–19.6). In 1996, half of the patients diagnosed as having bipolar or schizoaffective disorder were prescribed valproate. In the UK and other European countries, valproate is not even licensed for use in bipolar disorder and sales suggest that it is very little used. In the UK and Europe, the limited evidence available suggests that lithium remains the most commonly used mood stabiliser (Hill *et al*, 1996).

These findings suggest that valproate has been extremely effectively marketed in the USA. The increase in the prescription of valproate has occurred despite limited evidence for its effectiveness in the maintenance treatment of bipolar disorder (Emilien *et al*, 1996; Sharma *et al*, 1997).

The unavoidable consequence of these variations in clinical practice is considerable clinical uncertainty about the relative benefits of lithium and valproate. The most appropriate response to this uncertainty is to synthesise the current research evidence and to identify key clinical questions that require reliable answers. There are several Cochrane Systematic Reviews of treatments for bipolar disorder in progress:

- (a) lithium in maintenance treatment (Burgess *et al*, 2001);
- (b) valproate in acute treatment of mania (MacRitchie *et al*, 2001a);
- (c) valproate in maintenance treatment (MacRitchie *et al*, 2001b);

- (d) carbamazepine in maintenance treatment (Bandeira *et al*, 2001).

### Large-scale randomised evidence provides reliable answers to important clinical questions

The systematic reviews currently in progress suggest that existing trials are small and few in number, leaving considerable room for uncertainty about the best maintenance strategy in bipolar disorder. Finding answers to these important clinical questions requires large-scale, randomised evidence (Peto & Baigent, 1998). In other areas of medicine, large, simple, randomised trials have provided reliable answers to important clinical questions. For example, the ISIS-2 Collaborative Study demonstrated clearly that both aspirin and streptokinase prevented death following acute myocardial infarction and that the combination of both treatments was more effective than either alone (ISIS-2 Collaborative Group, 1988).

There are several characteristics of large, simple trials. First, to make them as relevant to as many future patients as possible, we need to recruit a heterogeneous group of patients (Peto *et al*, 1995). This is counter to the approach usually used for phase III trials, run primarily for regulatory purposes, that attempt to recruit a very narrowly defined group of patients. Rather than having restrictive entry criteria, the key entry criterion in large, simple trials is that there should be *substantial clinical uncertainty* about the best treatment for a particular patient. This approach may appear to miss potentially important differences between subgroups of patients, but this is usually not the case. *Quantitative* treatment interactions, in which a treatment is not as effective in one group as it is in another, are quite common in medicine. However, *qualitative* treatment interactions, in which a treatment is effective in one group and either not effective or actually harmful in another group, are rare.

Second, large, simple trials need to measure the comparative effect of the treatments on an outcome of direct clinical importance. Previous trials in psychiatry have often used primary outcomes of uncertain clinical meaning (Hotopf *et al*, 1997; Thornley & Adams, 1998).

Lastly, randomised, controlled trials need to be very large because most worthwhile treatment effects are only moderately sized and require large trials to measure

them reliably. Inadequate sample size is a widespread problem in drug trials in psychiatry (Johnson, 1983; Geddes *et al*, 1996; Hotopf *et al*, 1997; Thornley & Adams, 1998). Trials in general need to be much larger than they have been to date. Thousands of patients will often need to be randomised in the study to provide reliable and precise estimates. This requirement means that, to be feasible, large-scale clinical trials need to address a question of real and pressing clinical certainty and simplicity. If the question is not perceived as important, clinicians are likely to be less motivated to enter patients into the trial. The trials need to be kept simple to impose as little extra work on participating clinicians and patients as possible. Central administration can remove much of the administrative burden of trials.

### LARGE-SCALE RANDOMISED STUDIES IN BIPOLAR DISORDER

We believe there is sufficient clinical uncertainty about maintenance treatment in bipolar disorder to warrant a large, simple, randomised, controlled trial. The key clinical question is: in patients for whom prophylaxis has been recommended, which treatment most decreases the subsequent admission rate? The main uncertainty concerns the relative efficacy of lithium and sodium valproate (and the combination of the two drugs) in the maintenance treatment of bipolar disorder.

A large-scale, randomised study, the Bipolar Affective Disorder: Lithium Anticonvulsant Evaluation (BALANCE) study, is being planned to compare the combination of lithium and valproate with either drug alone. Here we discuss some of the key methodological issues involved in the design of this trial, which is now being piloted in several centres in the UK.

The main requirement is for reliable evidence on the comparative efficacy of the interventions. Although double-blinding might protect against performance and ascertainment biases, it would be difficult to achieve and the pilot study suggests that it would reduce participation. Ascertainment bias can be reduced by the use of a relatively "hard" primary outcome (see below). The primary question is about the relative efficacy of the treatments in patients who are willing to remain on long-term treatment, and so the trial needs to ensure that

the number of participants who drop out early is limited. There will therefore be a non-randomised run-in phase of 8 weeks. Participants who satisfactorily complete this run-in phase, and for whom there remains clinical uncertainty about the optimal maintenance treatment, will then be eligible for randomisation.

The primary outcomes in maintenance studies have been problematic in the past (Bowden *et al*, 1997). This is because of the complexity of the profile of disability caused by bipolar illness. In practice, the use of maintenance treatment with a mood stabiliser has the multiple aims of reducing the chances of suicide or severe recurrence requiring admission to hospital, and also of improving mood stability, reducing inter-episode mood symptoms and suicidal ideation, and improving overall quality of life. All these outcomes are important, but perhaps the most crucial is the reduction in the chances of severe relapse. We are therefore planning to use severe recurrence, defined by hospital admission, as the main outcome. The principal analysis will be survival analysis of time to readmission. Secondary outcomes will include clinical global impression, self-reported quality of life, adverse events, suicide attempts and use of additional psychotropic medication.

One consequence of using a clearly defined and clinically meaningful outcome such as hospital admission as the primary outcome is that a large sample will be required. An estimate of the probability of admission in bipolar disorder during a 2-year follow-up was provided by the Danish record linkage study (Kessing *et al*, 1998). It investigated the probability of readmission at different times in the course of the illness. Using these data, and assuming that most eligible patients would have had two to five recurrences, this produces an estimated expected readmission rate of approximately 50% over 2 years. Assuming an admission rate of 50% over 2 years and considering a 10% difference between treatments as clinically worthwhile, assuming a 30% drop-out rate (based on the mean drop-out rate from previous trials of maintenance treatment in bipolar disorder) for an  $\alpha=0.05$  and  $\beta=90\%$ , each group would need to include 713 patients – a total of 2139 in a three-arm study. It is possible that the event rates will be lower because, for a number of reasons, patients in trials may be less likely to relapse (Bowden *et al*, 1997). If the event rates were 30% on lithium and 20% on valproate, as observed

**Table 1** Required sample size per arm for various admission rates (assuming a 30% drop-out rate)

Admission rate (lithium)	Admission rate (valproate)							
	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0.20	1353	—	1746	483	236	145	101	76
0.25	391	1746	—	2092	565	271	162	111
0.30	196	483	2092	—	2381	631	296	175
0.40	88	145	271	631	2608	—	2769	713
0.50	55	76	111	175	315	713	2869	—

in the German carbamazepine study (Greil *et al*, 1997), 483 subjects would be required in each group. The sample sizes for various possible admission rates are shown in Table 1. Taking these estimates into account, we are aiming to recruit 3000 patients into BALANCE.

Non-randomised studies have suggested a potentially important specific anti-suicide effect for lithium. A review of studies in this area found that the average suicide rate in the studies was 3.2 per 100 patient-years for patients not on lithium, compared with 0.37 per 100 patient-years for patients taking lithium (Tondo *et al*, 1997). A study of 3000 patients would have over 90% power to detect such a treatment effect. At least some of the apparent anti-suicidal effect of lithium is due to selection bias in the non-randomised studies (patients with a better prognosis are more likely to take lithium); the true effect is likely to be lower and it is doubtful if the study would have sufficient power to reliably detect a more realistic effect. However, it is clearly important to try to obtain some measure of suicidality, and BALANCE will therefore measure the number of suicide attempts in the lithium groups.

Finally, we must mention logistics. We have now spoken about BALANCE to many clinicians with busy practices in the UK and have been impressed by the immediate positive response that has been the rule so far. Continued support from the medical community will be essential for the BALANCE study. The challenge is to devise a treatment choice that expresses the dilemma many clinicians feel – and all, perhaps, ought to feel, given the facts we have reviewed above. The biggest step, however, will be to randomise patients between the alternatives. In principle this is little different from the private randomisations we all currently make when we have no real basis for informed choice of

one option over another. If we can make it easy enough in practice to pool that uncertainty by a common act of randomisation, we can both solve the clinician's immediate dilemma and move on definitively to resolve it.

The BALANCE study may represent an important opportunity to define a new direction for clinical trials in psychiatry. Its development will carry lessons relevant to the evaluation of all drugs whose use is established in short-term studies but whose primary impact is intended to be in long-term prevention. By addressing a real clinical choice, it may prevent decisions being unduly influenced by simple marketing factors: it would be unacceptable if lithium use was increasingly discontinued because valproate is marketed actively and lithium is not. However, it is equally possible that undue conservatism in Europe is reducing the access of patients with bipolar disorder to a real alternative to lithium. We need to know which interpretation is correct. If successful, BALANCE could establish the pattern for maintenance trials of new drugs for long-term use in bipolar disorder.

## ACKNOWLEDGEMENTS

The pilot phase of BALANCE is funded by the Theodore and Vada Stanley Research Foundation.

## REFERENCES

- Bandeira, C. A., Lima, M. S., Geddes, J., *et al* (2001) Carbamazepine for bipolar affective disorders (protocol). *Cochrane Database of Systematic Reviews*, in press.
- Bowden, C. L., Swann, A. C., Calabrese, J. R., *et al* (1997) Maintenance clinical trials in bipolar disorder: design implications of the divalproex-lithium-placebo study. *Psychopharmacology Bulletin*, **33**, 693–699.
- , Calabrese, J. R., McElroy, S. L., *et al* (2000) A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. *Archives of General Psychiatry*, **57**, 481–489.

**Burgess, S., Geddes, J., Townsend, E., et al (2001)** Lithium for maintenance treatment of mood disorders (protocol). *Cochrane Database of Systematic Reviews*, in press.

**Citrome, L., Levine, J. & Allingham, B. (1998)** Utilization of valproate: extent of inpatient use in the New York State Office of Mental Health. *Psychiatric Quarterly*, **69**, 283–300.

**Emilien, G., Maloteaux, J. M., Seghers, A., et al (1996)** Lithium compared to valproic acid and carbamazepine in the treatment of mania: a statistical meta-analysis. *European Neuropsychopharmacology*, **6**, 245–252.

**Fenn, H. H., Robinson, D., Luvy, V., et al (1996)** Trends in pharmacotherapy of schizoaffective and bipolar affective disorders: a 5-year naturalistic study. *American Journal of Psychiatry*, **153**, 711–713.

**Freeman, H. (1999)** Psychiatry in the National Health Service 1948–1998. *British Journal of Psychiatry*, **175**, 3–11.

**Geddes, J. R. (1999)** Asking structured and focused clinical questions: essential first step of evidence-based practice. *Evidence-Based Mental Health*, **2**, 35–36.

— & **Harrison, P. J. (1997)** Closing the gap between research and practice. *British Journal of Psychiatry*, **171**, 220–225.

—, **Game, D., Jenkins, N. E., et al (1996)** What proportion of primary psychiatric interventions are based on randomised evidence? *Quality in Health Care*, **5**, 215–217.

—, **Wilczynski, N., Reynolds, S., et al (1999)** Evidence-based mental health – the first year. *Evidence-Based Mental Health*, **2**, 3–5.

**Greil, W., Ludwig Mayerhofer, W., Erazo, N., et al (1997)** Lithium versus carbamazepine in the maintenance treatment of bipolar disorders – a randomised study. *Journal of Affective Disorders*, **43**, 151–161.

**Hill, R. G., Hardy, P. & Shepherd, G. (1996)** *Perspectives on Manic Depression*. London: Sainsbury Centre for Mental Health.

**Hotopf, M., Lewis, G. & Normand, C. (1997)** Putting trials on trial – the costs and consequences of small trials in depression: a systematic review of methodology. *Journal of Epidemiology and Community Health*, **51**, 354–358.

**ISIS-2 (Second International Study of Infarct Survival) Collaborative Group (1988)** Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet*, **2**, 349–360.

**Jadad, A. R., Cook, D. J., Jones, A., et al (1998)** Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *Journal of the American Medical Association*, **280**, 278–280.

**Johnson, A. L. (1983)** Clinical trials in psychiatry. *Psychological Medicine*, **13**, 1–8.

**Kessing, L. V., Andersen, P. K., Mortensen, P. B., et al (1998)** Recurrence in affective disorder. I. Case register study. *British Journal of Psychiatry*, **172**, 23–28.

## CLINICAL IMPLICATIONS

- Evidence-based medicine can meet the information needs of clinicians by making full use of the best available evidence and identifying key areas of future research.
- The management of bipolar disorder is an important therapeutic area where opinion rather than evidence has been unduly influential.
- The Bipolar Affective Disorder: Lithium Anticonvulsant Evaluation (BALANCE) study is a large-scale, randomised, controlled trial comparing the long-term effectiveness of lithium and valproate alone or in combination in bipolar affective disorder.

## LIMITATIONS

- Many drug trials in psychiatry are too small and are done primarily for regulatory purposes.
- Existing trials are often of uncertain relevance to real-life clinical practice.
- Although large trials have been successfully completed in other areas of medicine, it is not yet known how feasible they may be in psychiatry.

JOHN GEDDES, MRCPsych, GUY GOODWIN, FRCPsych, Department of Psychiatry, University of Oxford, UK

Correspondence: Dr John Geddes, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK. Tel: +44 (0)1865 226480; fax: +44 (0)1865 793101; e-mail: john.geddes@psych.ox.ac.uk

**Manji, H. K., Moore, G. J. & Chen, G. (2001)** Bipolar disorder: leads from the molecular and cellular mechanisms of action of mood stabilisers. *British Journal of Psychiatry*, **178** (suppl. 41), s107–s119.

**MacRitchie, K., Geddes, J. R., Scott, J., et al (2001a)** Valproate for acute affective episodes in bipolar disorder (protocol). *Cochrane Database of Systematic Reviews*, in press.

—, —, —, et al (2001b) Valproate as maintenance treatment for bipolar affective disorder (protocol). *Cochrane Database of Systematic Reviews*, in press.

**Peto, R. & Baigent, C. (1998)** Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *British Medical Journal*, **317**, 1170–1171.

—, **Collins, R. & Gray, R. (1995)** Large-scale randomised evidence: large, simple trials and overviews of trials. *Journal of Clinical Epidemiology*, **48**, 23–40.

**Sackett, D. L., Haynes, R. B., Guyatt, G. H. & Tugwell, P. (1991)** *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston, MA: Little, Brown.

**Sanderson, D. R. (1998)** Use of mood stabilizers by hospitalized geriatric patients with bipolar disorder. *Psychiatric Services*, **49**, 1145–1147.

**Sharma, V., Yatham, L. N., Haslam, D. R., et al (1997)** Continuation and prophylactic treatment of bipolar disorder. *Canadian Journal of Psychiatry*, **42** (suppl. 2), 92S–100S.

**Stein, K. & Milne, R. (1999)** Mental health technology assessment: practice based research to support evidence-based practice. *Evidence-Based Mental Health*, **2** (2), 37–39.

**Thornley, B. & Adams, C. (1998)** Content and quality of 2000 controlled trials in schizophrenia over 50 years. *British Medical Journal*, **317**, 1181–1184.

**Tondo, L., Jamison, K. R. & Baldessarini, R. J. (1997)** Effect of lithium maintenance on suicidal behavior in major mood disorders. *Annals of the New York Academy of Sciences*, **836**, 339–351.