

Clinical trials of antidepressant medications are producing meaningless results

GORDON PARKER / IAN M. ANDERSON and PETER HADDAD

Edited and introduced by Mary Cannon, Kwame McKenzie and Andrew Sims.

INTRODUCTION

A recent alert from the UK Committee on Safety of Medicines stated that the dangers of treatment of depression with paroxetine outweigh the benefits in those under 18. Such a warning should focus our minds on the evidence on which clinical practice is based. Antidepressant treatment of depression in the under-18s has been thought to be justified because clinical trials show that it works so well in over-18s. But is that a reasonable assessment of the evidence? Kirsch *et al* (2002) use the analogy of 'The Emperor's New Clothes' to describe the findings from their meta-analysis of randomised placebo-controlled trials of antidepressants. They conclude that antidepressant medication appears to have only a small effect on outcome over and above placebo. In this analogy psychiatry is the emperor, drug trials are the fraudsters and the deception is being revealed by a growing body of critical opinion proposing that, once methodological problems with clinical trials are taken into account, antidepressants either do not work at all or have an effect that is so small as to be clinically unimportant (Andrews, 2001; Moncrieff, 2002). A large number of randomised placebo-controlled trials of antidepressants have been carried out over the past decades, mostly funded by the pharmaceutical industry, and it is now recognised that about 50% of negative trials go unpublished (Thase, 1999). Meanwhile, unipolar depression has jumped into the top five of the world's total burden of disease, and there is an imperative need for effective and safe treatments. Do we need more randomised controlled trials (RCTs) of antidepressant medications, or has that research paradigm outlived its usefulness? In this month's debate, Professor Gordon

Parker, University of New South Wales and Black Dog Institute, Australia, and Drs Ian Anderson and Peter Haddad from the University of Manchester discuss whether clinical trials for antidepressant medication produce meaningless results.

FOR

Two papers published last year add weight to an argument that efficacy data from clinical trials no longer provide meaningful evidence about the utility of antidepressant drugs.

The first (Hypericum Depression Trial Study Group, 2002) reported an 8-week double-blind, randomised placebo-controlled trial of St John's wort and the selective serotonin reuptake inhibitor sertraline. Despite substantive cell sizes, neither drug was significantly different from placebo in reducing depression severity or disability or in overall improvement. The second paper (Kirsch *et al*, 2002) analysed efficacy data submitted to the US Food and Drug Administration for the six 'most widely prescribed antidepressants' approved between 1987 and 1999. The mean drug-placebo difference (for 38 trials analysed) was two points on the Hamilton Rating Scale for Depression, allowing the authors to conclude that antidepressant drug effects were 'very small and of questionable clinical significance'. There is great difficulty reconciling such findings with clinical practice.

A related question is whether clinical trials provide any evidence that one type of antidepressant therapy is superior to any other. As reviewed earlier (Parker, 2001), very large databases suggest that different classes of antidepressant drugs are equally efficacious. Meta-analyses also report similar response rates for drugs and most non-drug treatments for depression. Robinson & Rickels (2002) reviewed about 60 psychotherapy studies and established only trivial superiority of pharmacotherapy

after controlling for the researcher's allegiance. The equipotency inference – that depression treatments are equally efficacious and not distinctly superior to placebo – is hard to reject from such evidence.

The equipotency theory has obvious implications. It fosters an 'affective fallacy' in both therapists and patients – evaluating therapies (particularly *their* therapy) impressionistically rather than by its integral strengths. High non-differential response rates allow therapists of many persuasions to claim their therapy as efficacious and scientifically proven.

Conversely, failure of therapies to differentiate from placebo invites a challenge that they act non-specifically, with Kirsch *et al* (2002) concluding, for example, that the 'pharmacological effects of antidepressants are clinically negligible'. Clinicians may view such conclusions as specious, but the public impact is not trivial. Patients benefiting from an antidepressant feel demeaned by media reports indicating that antidepressants are little better than placebos.

Several factors have contributed to such uninformative results. First, the current classificatory model. Rather than distinguishing separate depressive disorders (phenotypically or aetiologically), 'depression' is currently modelled as a single entity varying only in severity. Creating pseudo-entities such as 'major depression' for use as the principal 'diagnostic' measure increases the chance of non-differential results between interventions.

Second, recruitment procedures have led to unrepresentative trial subjects. Formal and informal screening excludes those with many comorbid conditions and the more 'biological' depressive disorders (e.g. melancholia). Inclusion and exclusion criteria ensure a pristine subject profile remote from clinical practice, and inviting redefinition of 'cosmetic psychopharmacology'.

Third, clinical trials remain subject to bias, despite the efforts of researchers and the use of placebo-controlled designs. High non-specific 'responsivity' of trial subjects is the fourth factor eroding their value. It is natural for humans to develop depressive reactions, which – for most – have the tendency to remit, whether spontaneously or in response to support or improvement in a stressful situation. Patients with 'clinical depression' differ by a distinctly lower likelihood of a 'spontaneous remission', whether reflecting biological, psychological or social factors. Subjects in clinical trials

are likely to be closer to the general community than to clinical patients in 'responsivity' terms – either to active treatment or to placebo – which is a more salient distorting factor than any 'placebo effect'. It is salutary to note that one analysis of controlled trials (Walsh *et al*, 2002) established a 7% per decade increase in the response rates to placebo and to antidepressant medication.

Thus, analyses such as that by Kirsch *et al* (2002) suggesting that response to the newer antidepressants only marginally exceeds placebo response is not surprising. In summary, current designs restrict the participation of 'true' specific responders, being overly weighted towards pristine subjects with non-biological depressive disorders, with unstable symptomatology and disorders of marginal severity, and disposed to 'respond' irrespective of the treatment arm. Extrapolation of such studies to the clinical management of melancholic depression, and possibly other 'biological' expressions of depression, is then illogical.

Should loss of confidence in trial data lead to their abandonment? Just as it is not necessary to abandon religion because of antipathy to the local minister, loss of faith in clinical trials might more usefully lead to modifying their faulty components. One strategy would be to reduce the distance between efficacy studies (assessing outcome under controlled conditions) and effectiveness studies (approximating to the clinical world), both by modifying the current efficacy study paradigm and by undertaking clinical panel studies.

It is hard to detect any winners from the current paradigm, whether licensing authorities, the pharmaceutical industry or patients. Current operational strategies for trials are producing specious and irrelevant information, compromising rationality and reality. They need to get real.

Declaration of interest

G.P. is a member of several international and Australian pharmaceutical advisory boards. This report is supported by an NHMRC Program Grant (222708) and a NSW Department of Health Infrastructure Grant.

Gordon Parker Scientia Professor, School of Psychiatry, University of New South Wales; Director, Black Dog Institute, Prince of Wales Hospital, Sydney, Australia. Correspondence: Euroa Unit, Prince of Wales Hospital, Randwick 2031, Australia.

AGAINST

Are drug trials of antidepressants more a triumph of marketing than science? Are RCTs a flawed way to test the efficacy of antidepressants? What if antidepressants do not really work, or at least not well enough to be important?

An increasingly vociferous minority is asserting just this (Antonuccio *et al*, 1999; Kirsch *et al*, 2002; Moncrieff, 2002). Irving Kirsch in particular has a populist appeal with titles such as 'Listening to Prozac but hearing placebo' (Kirsch & Sapirstein, 1998), and most recently 'The Emperor's new drugs' (Kirsch *et al*, 2002). It is uncomfortable to have our assumptions questioned. There is enormous investment in our belief that antidepressants work, from its buttressing the scientific basis of psychiatry, through our need as clinicians to have the tools for alleviating distress, to providing a financial return for pharmaceutical companies. However, as we psychiatrists know only too well, firmly held beliefs may, on occasion, be delusional. The fact that most people support the psychopharmacological orthodoxy is, in itself, no argument and it is important to examine the evidence before drawing conclusions.

Addressing the most forceful criticism, whether or not antidepressants really work (i.e. have a pharmacologically specific action), we do have to turn to RCT evidence as this is the only way to tease out the effects of placebo *v.* drug. The key issue is that of publication bias. Kirsch *et al* (2002) identified all available acute treatment studies comparing newer antidepressants with placebo in evidence submitted to the US Food and Drug Administration. This included previously unreported 'negative' studies and is likely to be as complete a data-set as possible. The outcome from pooling the studies is a highly statistically significant effect in favour of antidepressants. Therefore, whatever the size and cause of the effect, the central question as to whether there really is an effect is answered in the affirmative.

What is the cause of this effect? Greenberg *et al* (1992) have attributed it to unblinding. In other words, the effect exists but is due to an enhanced placebo effect because patients and/or assessors can tell who is receiving the active drug. Unfortunately, this not an objection that it is possible to answer definitively. The pharmaceutical industry and trialists have

done themselves no favours here because the relatively simple process of checking and reporting the success of blinding is rarely done. Nevertheless, strong circumstantial evidence all points to lack of blinding not accounting for the effect (Moncrieff *et al*, 1998; Smith *et al*, 2002; Geddes *et al*, 2003).

So, at least the Emperor's clothes are made of 'real cloth' (Thase, 2002). Nevertheless, we still have to answer the criticism that the fabric is so 'see-through' that it makes little difference – in other words, the weaker argument that the antidepressant effect is so small as to be clinically unimportant. Kirsch *et al* (2002) found that 80% of the effect of antidepressants was duplicated by placebo, a difference in end-point of about two Hamilton Rating Scale for Depression points, half the size of the effect usually reported (e.g. Anderson *et al*, 2000). For most patients, it can be argued, this is a marginal effect. One approach at this point is to enter a discussion about how big an effect is required to be clinically important, or to propose differential effects related to patient subgroups or depression severity. We believe this fundamentally misses the point; RCTs with soft end-points do not allow us to determine the size of effect in usual clinical practice. This is something that evidence-based medicine, at least in its current guise, has misled us about.

Quantification, numbers needed to treat, effect sizes extrapolated to clinical practice are all based on assumptions that are violated in RCTs of antidepressants. First, the assumption that RCT patients are representative of the clinical population is simply untrue (Zimmerman & Posternak, 2002). Second, the overall effect size for antidepressants may be crucially dependent on the size of the placebo effect. This is most evident if change from baseline is used as the yardstick as in Kirsch *et al* (2002).

Our position is, therefore, that we simply do not know how big the effect of antidepressants is in clinical practice because RCTs are not designed to tell us this. Clinical trials of antidepressants are not producing meaningless results, because they can tell us which compounds work (i.e. have efficacy). This is vitally important both scientifically and as a cornerstone of the regulatory process, designed to ensure that drugs that are licensed are safe and have a real effect. What is meaningless is to ask the trials questions they cannot answer, such as how well do antidepressants

work in usual practice (their effectiveness). The latter question needs different trial designs from that of the standard RCT. This is no easy task and is one that will require more pragmatic/naturalistic approaches to be more inclusive, while attempting to minimise allocation bias. There needs to be careful selection of target groups, comparison treatments and duration of the assessment period. Only then will we be able to estimate the real added value of antidepressants in particular patient groups.

Declaration of interest

I.M.A. and P.H. have both received honoraria for speaking, been members of advisory boards, received research grants and had support for attending scientific meetings from several pharmaceutical companies involved in the manufacture and marketing of antidepressants.

Ian M. Anderson Senior Lecturer in Psychiatry, Neuroscience and Psychiatry Unit, University of Manchester, Room G809 Stopford Building, Oxford Road, Manchester M13 9WL, UK. Tel: 0161 275 7428; Fax: 0161 275 7429; e-mail: ian.anderson@man.ac.uk

Peter Haddad Consultant Psychiatrist, Mental Health Partnership of Bolton, Salford & Trafford, Manchester, UK

REFERENCES

- Anderson, I. M., Nutt, D. J. & Deakin, J. F. W. (2000)** Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 1993 British Association for Psychopharmacology guidelines. *Journal of Psychopharmacology*, **14**, 3–20.
- Andrews, G. (2001)** Placebo response in depression: bane of research, boon to therapy. *British Journal of Psychiatry*, **178**, 192–194.
- Antonuccio, D. O., Danton, W. G., DeNelsky, G. Y., et al (1999)** Raising questions about antidepressants. *Psychotherapy and Psychosomatics*, **68**, 3–14.
- Geddes, J. R., Carney, S. M., Davies, C., et al (2003)** Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet*, **361**, 653–661.
- Greenberg, R. P., Bornstein, R. F., Greenberg, M. D., et al (1992)** A meta-analysis of antidepressant outcome under 'blinder' conditions. *Journal of Consulting and Clinical Psychology*, **60**, 664–669.
- Hypericum Depression Trial Study Group (2002)** Effect of *Hypericum perforatum* (St John's wort) in major depressive disorder: a randomised controlled trial. *Journal of the American Medical Association*, **287**, 1807–1814.
- Kirsch, I. & Sapirstein, G. (1998)** Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication. *Prevention & Treatment*, **1**, Article 2. <http://www.journals.apa.org/prevention/volume1/pre0010002a.html>
- , Moore, T. J., Scoboria, A., et al (2002)** The Emperor's new drugs: An analysis of antidepressant medication data submitted to the US Food and Drug Administration. *Prevention & Treatment*, **5**, Article 23. <http://www.journals.apa.org/prevention/volume5/pre0050023a.html>
- Moncrieff, J. (2002)** The antidepressant debate. *British Journal of Psychiatry*, **180**, 193–194.
- , Wessely, S. & Hardy, R. (1998)** Meta-analysis of trials comparing antidepressants with active placebos. *British Journal of Psychiatry*, **172**, 227–231.
- Parker, G. (2001)** 'New' and 'old' antidepressants: all equal in the eyes of the lore? *British Journal of Psychiatry*, **179**, 95–96.
- Robinson, D. S. & Rickels, K. (2002)** Concerns about clinical drug trials. *Journal of Clinical Psychopharmacology*, **20**, 593–596.
- Smith, D., Dempster, C., Glanville, J., et al (2002)** Efficacy and tolerability of venlafaxine compared with selective serotonin reuptake inhibitors and other antidepressants: a meta-analysis. *British Journal of Psychiatry*, **180**, 396–404.
- Thase, M. (1999)** How should efficacy be evaluated in randomised controlled trials of treatments for depression? *Journal of Clinical Psychiatry*, **60** (suppl. 4), 23–31.
- (2002)** Antidepressant effects: the suit may be small, but the fabric is real. *Prevention & Treatment*, **5**, Article 32. <http://www.journals.apa.org/prevention/volume5/pre0050032c.html>
- Walsh, B. T., Seidman, S. N., Sysko, R., et al (2002)** Placebo response in studies of major depression: variable, substantial, and growing. *Journal of the American Medical Association*, **287**, 1840–1847.
- Zimmerman, M. & Posternak, M. A. (2002)** Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *American Journal of Psychiatry*, **159**, 469–473.