Editorial

Antidepressants on trial: how valid is the evidence?

Gordon Parker

Summary

A recent meta-analysis concluded that newer antidepressant drugs are equivalent to or no better than placebos, a conclusion at some variance with their commonly judged clinical effectiveness. The 'disconnect' between randomised controlled trials and clinical practice would benefit from dissection of contributing factors, and redressing limitations to current trial procedures.

Declaration of interest

G.P. serves on several advisory boards for antidepressant drugs, and has chaired and spoken at meetings convened by pharmaceutical companies.

Gordon Parker (pictured) is Scientia Professor of Psychiatry at the University of New South Wales and Executive Director of the Black Dog Institute in Sydney, Australia.

'There is no treatment they cannot make equal to placebo.'[1]

A recent meta-analytic study of randomised controlled trial (RCT) data by Kirsch et al[2] effectively concluded that the new antidepressant drugs are either no better than placebos or only as effective as placebos, generating ingenuous acceptance in many lay and medical publications, and dismissal by many clinicians who view antidepressant drugs as highly effective. The risk is of debating the findings rather than the constituent processes. This editorial considers why findings from RCT evidential bases should not be viewed as usefully generalising to clinical application.

Do RCTs identify differential effectiveness across treatments of major depression?

In the past two decades the efficacy of most antidepressant treatments has generally been tested in relation to major depression. However, evidence of any 'treatment specificity effects' is hard to find, despite enormous databases. As overviewed elsewhere,[3] meta-analyses comparing (a) 'old' (e.g. tricyclics) and 'new' antidepressant drugs (e.g. selective serotonin reuptake inhibitors), (b) differing psychotherapies, and (c) pharmacotherapy v. psychotherapy, return comparable efficacy rates. All evaluated treatments appear equally efficacious for 'major depression'.

Do antidepressant treatments differentiate from placebos in RCTs?

Moncrieff et al[4] undertook a meta-analysis comparing tricyclic antidepressants with placebos – and as only two of the nine studies favoured the drug, the authors argued for similar meta-analyses of the newer antidepressants. In 2002, Kirsch et al[5] published such an analysis, examining RCT data for six new antidepressants. Of the 47 data-sets, the antidepressant did not differentiate from placebo in 9, and for the remaining 38, the drug–placebo difference was a 'trivial' two points. Their more recent report[2] analysed data of 35 RCTs comparing 5133 participants randomised to medication and 1841 to placebo, with weighted mean improvement in depression severity being 9.6 and 7.8 points respectively, but with baseline depression severity influencing drug efficacy. The authors concluded that 'the overall effect of new generation antidepressant medications is below recommended criteria for clinical significance' and that 'there seems little evidence to support the prescription of antidepressant medication to any but the most severely depressed patients, unless alternative treatments have failed to provide benefit'.

In relation to the latter, it is often unappreciated that the so-called evidence-supported psychotherapies (i.e. cognitive–behavioural therapy and interpersonal psychotherapy) also show non-differentiation from plausible control strategies in similar meta-analyses.[6]

Why the disconnect between RCTs and clinical practice?

Such findings allow three possible explanations: the therapies are ineffective, analyses are inappropriate or limitations to RCT procedures.

The first explanation is relative, not absolute – it is unlikely that the 'evidence-based' antidepressant therapies are always ineffective. The second explanation (effectively, 'garbage in, garbage out') was well-addressed by Lieberman et al,[7] who detailed problems from conducting, reporting and evaluating meta-analyses involving intent-to-treat and last-observation-carried-forward strategies, differential attrition, drug dosing (flexible v. fixed), participant sampling and 'cherry-picking' rather than including all relevant studies. The third explanation – that there are substantive limitations to current procedures for testing antidepressant treatments – is argued here as the most sustainable.

Contribution of the criterion diagnosis of major depression

Imagine if major dyspnoea was the criterion diagnosis for an RCT comparing a putatively effective treatment and a placebo. Further assume that study participants had various respiratory conditions (pneumonia, asthma, pulmonary embolus). It would be illogical to test a specific treatment (e.g. antibiotic, bronchodilator, anti-coagulant) as if it had universal application as results would be influenced by the prevalence of the constituent pathological disorders. A truly effective treatment would have its efficacy diminished or nullified by low representation of the target condition.

Thus, if major depression is no more than a 'domain diagnosis' – encapsulating differing constituent disorders (variably responsive to medication or to a psychotherapy) – then the true

efficacy of each treatment modality is at risk of clouding. Viewing major depression as a unitary entity – as against a non-specific domain diagnosis capturing heterogeneous expressions of depression – is a starting point for downstream non-specific results.

Impact of participant definition in RCTs

Most antidepressant drug trials recruit out-patients and effectively exclude those with melancholic depression – the quintessential 'biological' depressive condition. Also excluded are those with suicidal ideation, comorbid drug or alcohol problems, anxiety conditions and/or personality disorders. Individuals are commonly recruited via public advertising and may be reimbursed, and trial incentives risk rating up those with less substantive disorders to meet entry criteria. Such criteria risk recruiting individuals with less severe non-melancholic disorders and showing little correspondence with depressed patients presenting to psychiatrists.

As detailed by Lieberman et al,[7] early RCTs of antidepressants were weighted to hospitalised patients and those with the more biological mood disorders, with drug–placebo differences of 30%. As recruitment is increasingly weighted to those with milder, briefer and self-limiting expressions of depression – with Walsh et al[8] quantifying a 7% per decade increase in RCT responder rates for antidepressant drug and placebo – the increased spontaneous remission rates compromise detecting any signal from truly efficacious antidepressant drugs.

Influence of depression severity in RCTs

Horowitz & Wakefield[9] have detailed the risk of DSM-defined major depression pathologising states of normal sadness – and the 'myth of equivalence' (of equating symptom-based diagnoses across community and clinical samples).

At some decreasing level of severity, antidepressant drug treatments may move from being effective to ineffective – as quantified in the recent meta-analysis[2] – purely reflecting severity or reflecting low prevalence of the more severe biological conditions more specifically responsive to antidepressant drugs.

Further, severity-based measures risk being problematic at lower severity levels. First, some individuals (including those who might benefit from medication) may not yet be at the nadir of their illness. As a consequence, the true impact of an intervention might be compromised at that time. Second is the difficulty of separating state depression from base functioning. In clinical practice, an optimal target is for the patient to feel 'back to normal'. However, 'normality' might include (say) some distractibility, sleep and appetite disturbance – all symptoms that generate scores on state depression measures. Thus, non-remission status in an RCT might reflect a truly ineffective treatment, a partially effective treatment or merely general functioning.

In most RCTs, however, the primary outcome measure is 'responder' status. Baseline inflations for recruitment purposes,[7] together with individuals' placebo and spontaneous improvement propensities, risk regression to the mean confounding responder status. Responder status may be achieved by true- and false-positive improvers.

Thus, RCTs risk imprecision if outcome is responder status and confounding by trait functioning if outcome is remission status. Although corrective analytic strategies (e.g. mixed model repeated measures) have been suggested,[7] these are rarely adopted.

Alternative non-severity models for defining samples for treatment evaluation

In the absence of distinct biological markers, psychiatry used to weigh phenomenological strategies defining clinical phenotypes and/or causal factors.

Any reprised phenomenological model should prioritise psychotic and melancholic depression as candidate conditions for demonstrating selective and distinctive response to antidepressant drugs. As reviewed elsewhere,[3] studies in the 1960s – in which antidepressants differentiated distinctly from placebos – were weighted to the melancholic depressive subtype, and generated response rates of 60–70% to broad-spectrum antidepressant drugs, with placebo rates as low as 10%.

McHugh[10] has argued for four aetiopathic clusters of mental disorders, including clusters comprising 'patients with brain diseases' (e.g. psychotic and melancholic depressions), weighting causal factors emerging from temperament or personality level, and conditions provoked by significant life events.

Antidepressant drugs might be superior for those in the first group; psychotherapies (e.g. cognitive–behavioural therapy) correcting causal personality factors might be more salient for those in the second group; and interpersonal psychotherapy and counselling might be more effective for those in the third group. The argument put here is that no treatment should be viewed (or trialled) as having universal (or non-specific) application across heterogeneous disorders. Rather than selecting treatments on such as basis – or for eclectic reasons – the field would benefit from a model that specifies treatments weighted to differing biological, psychological and social factors contributing to depressive patterns.

Conclusions

If we are to argue that antidepressant drugs are evidence based, then we need to reconcile the reality that the largest referenced databases provide limited support for that proposition. The meta-analyses by Kirsch et al[2,5] principally analysed data used by pharmaceutical companies to argue the efficacy of antidepressant drugs (and have them licensed). If we wish to reject the imputation that antidepressant drugs are little better than placebos, we need first to recognise limitations of current RCT procedures and produce better evidence.

The position put here is not to reject the necessity for RCTs to begin to inform us about efficacy (and safety) of antidepressant drugs, but to argue that the limited findings should drive concerns about current diagnostic classifications, RCT procedures (whereby the 'apples' assessed in such studies do not correspond to the 'oranges' of clinical practice), reliance of treatment guidelines on such RCT findings and how the evidence-based depression treatments have been positioned at the expense of appropriate explanatory models. Trialling a (drug or non-drug) treatment as if it had universal (i.e. non-specific) application for a non-specific condition (e.g. major depression) risks building to non-specific results. The Kirsch meta-analysis[2] informs us that the consequences of such flawed logic have now been realised. The current foundations lack a firm base, and the meta-analysis has exposed a fault line, with flawed paradigms and RCT practices generating limited valid evidence.

Gordon Parker, School of Psychiatry, University of New South Wales, Black Dog Institute, Randwick 2031, Sydney, Australia. Email: g.parker@unsw.edu.au

References

1 Himmelhoch JM. On the usefulness of clinical case studies (comment). *Bipolar Disord* 2003; **5**: 69–71.

2 Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008; **5**: e45.

3 Parker G, Manicavasagar V. *Modelling and Managing the Depressive Disorders: A Clinical Guide.* Cambridge University Press, 2005.

4 Moncrieff J, Wessely S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebos. *Br J Psychiatry* 1998; **172**: 227–31.

5 Kirsch I, Moore T, Scoboria A, Nicholls SS. The emperor's new drugs. An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prev Treat* 2002; **5**: 1–11.

6 Parker G, Fletcher K. Treating depression with the evidence-based psychotherapies: a critique of the evidence. *Acta Psychiatr Scand* 2007; **115**: 352–9.

7 Lieberman JA, Greenhouse J, Hamer RM, Krishnan KR, Nemeroff CB, Sheehan DV, et al. Comparing the effects of antidepressants: consensus guidelines for evaluating quantitative reviews of antidepressant efficacy. *Neuropsychopharmacology* 2005; **30**: 445–60.

8 Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; **287**: 1840–7.

9 Horowitz AV, Wakefield JC. *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder.* Oxford University Press, 2007.

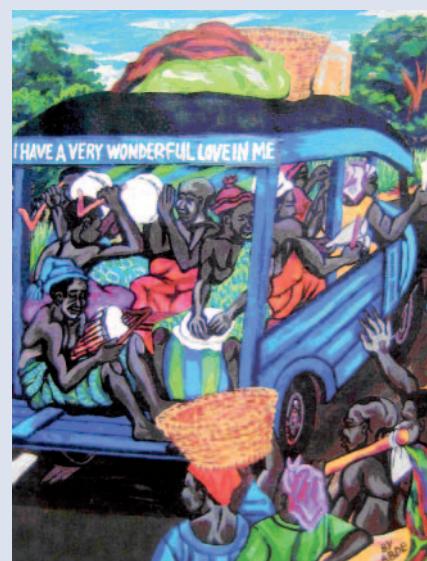10 McHugh PR. Striving for coherence: psychiatry's efforts over classification. *JAMA*. 2005; **293**: 2526–8.

psychiatry in pictures

I have a very wonderful love in me
by Abdullah JoJo Peter

The author of *I have a very wonderful love in me* is by Mr Abdullah JoJo Peter, a talented artist working in Accra, Ghana. Mr Abdullah JoJo Peter has schizophrenia. However, with the right treatment and care from Accra Mental Hospital he is recovering from his illness.

The image has been used as the cover for a book *Essential Skills for Mental Health Care* that is being distributed free of charge to health workers in Africa. The book has been developed with the help of the NGO BasicNeeds (www.basicneeds.org) in conjunction with the Ghana Health Service. It is free from copyright restrictions and may be downloaded from the BasicNeeds website (http://www.basicneeds.org/ghana/ghana2.asp).

Researched by Jim Crabb