



The limitations of using randomised controlled trials as a basis for developing treatment guidelines

Roger Mulder,^{1,2} Ajeet B Singh,^{1,3} Amber Hamilton,^{1,4,5,6} Pritha Das,^{1,4,5,6} Tim Outhred,^{1,4,5,6} Grace Morris,^{1,4,5,6} Darryl Bassett,^{1,7} Bernhard T Baune,^{1,8} Michael Berk,^{1,3,9} Philip Boyce,^{1,10} Bill Lyndon,^{1,5,11,12} Gordon Parker,^{1,13,14} Gin S Malhi^{1,4,6}

¹Mood Assessment and Classification (MAC) Committee, Sydney, Australia; ²Department of Psychological Medicine, University of Otago, Christchurch, New Zealand; ³School of Medicine, IMPACT Strategic Research Centre, Deakin University, Barwon Health, Geelong, Victoria, Australia; ⁴Department of Psychiatry, Northern Sydney Local Health District, St Leonards, New South Wales, Australia; ⁵Sydney Medical School Northern, University of Sydney, Sydney, New South Wales, Australia; ⁶CADE Clinic, Royal North Shore Hospital, Northern Sydney Local Health District, St Leonards, New South Wales, Australia; ⁷Private Practice in Psychiatry and Division of Psychiatry, University of Western Australia, Perth, Australia; ⁸Discipline of Psychiatry, University of Adelaide, Adelaide, South Australia, Australia; ⁹Department of Psychiatry, Orygen Research Centre, and the Florey Institute for Neuroscience and Mental Health, University of Melbourne, Melbourne, Victoria, Australia; ¹⁰Discipline of Psychiatry, Sydney Medical School, Westmead Clinical School, University of Sydney, Sydney, New South Wales, Australia; ¹¹Mood Disorders Unit, Northside Clinic, Greenwich, New South Wales, Australia; ¹²ECT Services, Northside Group Hospitals, Greenwich, New South Wales, Australia; ¹³School of Psychiatry, University of New South Wales, Kensington, New South Wales, Australia; ¹⁴Black Dog Institute, Sydney, New South Wales, Australia

Correspondence to Professor Gin S Malhi, CADE Clinic, Academic Department of Psychiatry, Level 3, Main Building, Royal North Shore Hospital, St Leonards, Sydney, NSW 2065, Australia; gin.malhi@sydney.edu.au

ABSTRACT

Randomised controlled trials (RCTs) are considered the ‘gold standard’ by which novel psychotropic medications and psychological interventions are evaluated and consequently adopted into widespread clinical practice. However, there are some limitations to using RCTs as the basis for developing treatment guidelines. While RCTs allow researchers to determine whether a given medication or intervention is effective in a specific patient sample, for practicing clinicians it is more important to know whether it will work for their particular patient in their particular setting. This information cannot be garnered from an RCT. These inherent limitations are exacerbated by biases in design, recruitment, sample populations and data analysis that are inevitable in real-world studies. While trial registration and CONSORT have been implemented to correct and improve these issues, it is worrying that many trials fail to achieve such standards and yet their findings are used to inform clinical decision making. This perspective piece questions the assumptions of RCTs and highlights the widespread distortion of findings that currently undermine the credibility of this powerful design. It is recommended that the clinical guidelines include advice as to what should be considered good and relevant evidence and that external bodies continue to monitor RCTs to ensure that the outcomes published indeed reflect reality.

In preparing the Royal Australian and New Zealand College of Psychiatrists guidelines for mood disorders,¹ the usual empirical methodological hierarchy was employed in which individual case reports were at the ‘bottom’ and randomised controlled trials (RCTs) at the ‘top’. This action is virtually unthinking, reflecting the rise of evidenced-based medicine. What is good about RCTs? The canonical answer is that RCTs control for unknown confounders by a design that ensures that all features causally related to the outcome other than treatment are distributed identically between the treatment and control groups. If the outcome is more probable in the treatment group, then the only explanation possible is that the treatment caused the outcome in some members of that group.² But as Cartwright has pointed out, the logic of RCTs is ideal for supporting ‘it-works-somewhere’ claims. Demonstrating that a drug is effective in a patient sample is an essential step in drug registration, hence the need for clinical trials to be conducted so that the drug can get to market. In clinical practice, we need evidence for its clinical utility; that it will produce the desired outcome in real-world patients and settings: the ‘it-works-for-us’ claim. This article questions the truth of both claims in the context of RCTs informing mood disorder clinical guidelines.

‘IT-WORKS-SOMEWHERE’

First, the ‘it-works-somewhere’ claim will be evaluated. RCTs in psychiatry may have bias in design, recruitment, patient populations, data analysis and presentation of findings. Studies are relatively small generally involving, at most, a few hundred subjects. The treatment effect sizes are small, which compounds the problem of clinical utility and translation. Many syndromes have high spontaneous recovery and placebo response rates, which complicate analyses and obfuscate

effects. Definitions of syndromes are often imprecise and overlapping while simultaneously heterogeneous, and commonly result in highly mixed samples. Added to this, a variety of outcome measures are used while interventions are often only conducted in tertiary referral units.³ Such specialist units are usually found in academic institutions, where patients are referred due to more complex illness patterns. Such groups of patients tend to have poorer prospects of remission and are rarely reflective of the general population with mood disorders.

There is consistent evidence of selected or distorted reporting in RCTs. Chan *et al*⁴ reviewed 102 trials and noted that the reporting of trial outcomes is not only frequently incomplete but also biased and inconsistent with protocols. Over half the trials were reported either in part (incompletely) or not at all, with statistically significant results having higher odds of being reported compared with non-significant outcomes for both efficacy (pooled OR 2.4) and harm (pooled OR 4.7). In other words, a significant outcome results in a greater likelihood of being reported. More disturbingly, 86% of survey responders denied the existence of unreported outcomes despite evidence to the contrary. A prominent example of this in psychiatry is Study 329: an RCT comparing paroxetine, imipramine and placebo in adolescents. A recent re-analysis reported that paroxetine only produced a positive result when four new secondary outcome measures were used instead of the primary outcomes. Analysing the primary outcome measure revealed no group differences.⁵

Due to these concerns, the International Committee of Medical Journal Editors introduced a policy to require registration of all clinical trials prior to enrolment of subjects.⁶ Registration involves information about trial protocols and the specified outcome measures being made publicly available. However, the efficacy of trial registration has been called into question by

a number of studies. In a review of five psychiatry journals that mandate registering of prospective clinical trials, it was reported that only 33% of trials were correctly prospectively registered and, of these, 28% had evidence of selective outcome reporting and 27% a large change in participant numbers. Overall, only 14.4% were correctly registered and reported.⁷ For psychotherapy RCTs, the results were even worse. Only 24.1% were registered and 4.5% free from selective outcome reporting,⁸ underscoring the fact that bias is not an issue just confined to pharmaceutical industry trials.

The other major attempt to improve the conduct and reporting of RCTs is the Consolidated Standards of Report Trials (CONSORT) guidelines. These provide an evidence-based minimum set of recommendations for reporting RCTs.⁹ While there is evidence that reporting in psychiatric RCTs has improved, over 40% of studies still do not adhere to the CONSORT guidelines.¹⁰

There is also the often cited influence of pharmaceutical marketing where there may be motivation for bias in design and external validity. A review of drug company authorship and sponsorship on drug trial outcomes reported that of 198 studies in three prestigious psychiatry journals (*British Journal of Psychiatry*, *American Journal of Psychiatry* and *JAMA Psychiatry*), only 23% were independently funded. Furthermore, independently funded studies were significantly more likely to report negative findings whereas industry-authored studies nearly always reported positive findings. Specifically, 74 out of 76 RCTs in this study demonstrated this bias¹¹—although journal editors are also reluctant to publish negative studies suggesting that sources of this bias arise at multiple levels. Similar effects are found in psychotherapy RCTs. Larger positive effect sizes were found when authors' allegiance to the studied psychotherapy existed and this allegiance effect was even stronger where the RCT was performed by the developer of the preferred treatment.^{12 13} Finally, the influence of publication bias remains an issue. A large survey of RCT researchers (n=318) revealed that around 25% of trials go unpublished and these unpublished studies are less likely to have favoured the new therapy. Interestingly, they noted that non-publication was primarily a result of failure to write up and submit trial results rather than a rejection of submitted manuscripts.¹⁴

Overall, while the 'it-works-somewhere' claim is somewhat more likely to be true over the past decade, it remains possible that many published RCTs are spurious or at least overstate their claims through a combination of methodological flaws (eg, type I and type II errors), selective reporting, marketing interests and publication bias.

'IT-WILL-WORK-FOR-US'

The 'it-will-work-for-us' claim is more difficult to evaluate. The CONSORT statement mandates a clear exposition of the recruitment pathway by which patients enter the RCT. This reporting is intended to enable clinicians to judge to whom the results of the RCT apply. But again the reality falls short. A review of trials leading to clinical alerts by the US National Institute of Health revealed that in relation to 31 eligibility criteria, only 63% were published in the main trial report and only 19% in the clinical alert.¹⁵ Inadequate reporting is even more of a problem in secondary publications such as clinical guidelines since space limitations and the need for a succinct message do not allow for detailed consideration of eligibility of trials or other determinants of external validity.¹⁶ Exclusion of common comorbidities is one of the common factors preventing real-world generalisability of RCTs. Furthermore, the population-level statistical approach to evidence-based medicine can 'homogenise' the complex heterogeneity of clinical reality and produce empirical data sets that lack clinical salience to real-world patients. It is not good to be an outlier when a standard population-based evidence approach is applied to your care.

Much less discussed (other than the type II error bias)¹⁷ is the possibility that true findings may be annulled because of reverse bias (ie, bias may under-estimate treatment effect). A potential source of reverse bias

in psychiatric RCTs emerges from recruitment strategies. For example, participants entering clinical trials for depression are likely to be mild or moderately depressed, sometimes better diagnosed as having an adjustment disorder or a persistent depressive disorder, or a depression related to psychosocial adversity, but are all lumped into a 'major depression' category to meet recruitment targets. Patients may inflate their scores to get 'free' treatment while assessing clinicians may inflate scores to enhance recruitment.¹⁸ Many trials exclude those with common comorbidities, such as those with suicidal ideation, and only seldom is a developmental trauma history obtained to better inform diagnosis.¹⁹

CONCLUSION

While RCTs provide the most credible research design for evaluating the effects of treatment at population levels, there is justifiable concern that the way the trials are conducted results in limited external validity and clinical salience. Despite efforts using trial registration and CONSORT, the evidence indicates many RCTs fall short of these standards. Further bias is introduced by pharmaceutical industry funding, 'championing' by developers of psychotherapies and publication bias. Clinical practice guidelines leave judgement largely to clinicians governed by clinical experience, but their observations do carry weight and inform decisions regarding patient care. Although this may seem inadequate, it reflects the current lack of explicit methodology to evaluate efficacy claims down to the level of individual patient decision-making. It can be argued that clinical guidelines need to include advice about what counts as good and relevant evidence.²⁰ To use RCT evidence, we need to tackle rather than ignore the real issues of whether 'it-works-somewhere' is actually true and even more whether this means 'it-will-work-for-us'. Applying empirical evidence to effectively care for the individual patient is the 'art' of medicine. It is an art that is alive and well, but can lead to idiosyncratic practice, making guidelines still pertinent despite the many epistemological limitations of population-level clinical trial science.

Collaborators The Mood Assessment and Classification Committee (MAC Committee) comprised academic psychiatrists with clinical expertise in the management of mood disorders and researchers with an interest in depression and bipolar disorders. The independently convened committee specifically targeted contentious aspects of mood disorders diagnosis and assessment with the express aim of informing clinical practice and future research. Members of the committee held one face to face meeting in Sydney (Australia) to discuss the issues in depth and agree upon outcomes. These were then developed further via email correspondence.

Funding The MAC Project was supported logistically by Servier who provided financial assistance with travel and accommodation for those MAC Committee members travelling interstate or overseas to attend the meeting in Sydney (held on 18th March 2017). Members of the committee were not paid to participate in this project and Servier had no input into the content, format or outputs from this project.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

doi:10.1136/eb-2017-102701

Received 28 April 2017; Revised 26 May 2017; Accepted 29 May 2017

REFERENCES

1. **Malhi GS**, Bassett D, Boyce P, *et al*. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Aust N Z J Psychiatry* 2015;**49**:1087–206.
2. **Cartwright N**. A philosopher's view of the long road from RCTs to effectiveness. *Lancet* 2011;**377**:1400–1.
3. **Mulder RT**, Frampton C, Joyce PR, *et al*. Randomized controlled trials in psychiatry. Part II: their relationship to clinical practice. *Aust N Z J Psychiatry* 2003;**37**:265–9.
4. **Chan AW**, Hróbjartsson A, Haahr MT, *et al*. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;**291**:2457–65.
5. **Barber S**, Cipriani A. Lessons learned from Restoring Study 329: transparent reporting, open databases and network meta-analyses as the way forward. *Aust N Z J Psychiatry* 2017;**51**:407–9.

6. **De Angelis C**, Drazen JM, Frizelle FA, *et al*. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *Lancet* 2004;**364**:911–2.
7. **Scott A**, Rucklidge JJ, Mulder RT. Is mandatory prospective trial registration working to prevent publication of unregistered trials and selective outcome reporting? An observational study of five psychiatry Journals that mandate prospective clinical trial registration. *PLoS One* 2015;**10**:e0133718.
8. **Bradley HA**, Rucklidge JJ, Mulder RT. A systematic review of trial registration and selective outcome reporting in psychotherapy randomized controlled trials. *Acta Psychiatr Scand* 2017;**135**:65–77.
9. **Begg C**, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;**276**:637–9.
10. **Han C**, Kwak KP, Marks DM, *et al*. The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. *Contemp Clin Trials* 2009;**30**:116–22.
11. **Tungaraza T**, Poole R. Influence of drug company authorship and sponsorship on drug trial outcomes. *Br J Psychiatry* 2007;**191**:82–3.
12. **Dragioti E**, Dimoliatis I, Fountoulakis KN, *et al*. A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Ann Gen Psychiatry* 2015;**14**:25.
13. **Lundh LG**, Petersson T, Wolgast M. The neglect of treatment-construct validity in psychotherapy research: a systematic review of comparative RCTs of psychotherapy for borderline personality disorder. *BMC Psychol* 2016;**4**:44.
14. **Dickersin K**, Chan S, Chalmers TC, *et al*. Publication bias and clinical trials. *Control Clin Trials* 1987;**8**:343–53.
15. **Shapiro SH**, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epidemiol* 2000;**53**:973–9.
16. **Rothwell PM**. External validity of randomised controlled trials: “to whom do the results of this trial apply?”. *Lancet* 2005;**365**:82–93.
17. **Edlund MJ**, Overall JE, Rhoades HM. Beta, or type II error in psychiatric controlled clinical trials. *J Psychiatr Res* 1985;**19**:563–7.
18. **Geddes JR**, Cipriani A. Time to abandon placebo control in pivotal phase III trials? *World Psychiatry* 2015;**14**:306–7.
19. **Cipriani A**, Zhou X, Del Giovane C, *et al*. Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *Lancet* 2016;**388**:881–90.
20. **Leucht S**, Chaimani A, Cipriani AS, *et al*. Network meta-analyses should be the highest level of evidence in treatment guidelines. *Eur Arch Psychiatry Clin Neurosci* 2016;**266**:477–80.