


Poorly Recognized and Uncommonly Acknowledged Limitations of Randomized Controlled Trials

Chittaranjan Andrade¹ 

ABSTRACT

Randomized controlled trials (RCTs) are a gold standard in evidence-based research. However, RCTs have limitations, among which the most commonly acknowledged is that narrow study selection criteria compromise the external validity of the findings. This article briefly touches upon this and other well-recognized limitations and presents, in greater detail, less commonly acknowledged limitations with examples from contemporary literature. Important among the less commonly acknowledged limitations are biases in RCTs of interventions to which patients cannot be blinded, weaknesses in the design of maintenance therapy RCTs, and, ubiquitously, post-randomization biases. The listed limitations notwithstanding, RCTs are still the best among research designs. What is important is to recognize the imperfections in each RCT so that the findings of the RCT can be better judged.

Keywords: Randomized controlled trials, limitations, internal validity, external validity post-randomization bias, blinding

A good study should have internal validity; that is, it should provide an unbiased and trustworthy answer to the research question. A good study should also have external validity; that is, its findings should be true not just for the study sample but for the population to which the findings are expected to be applied.¹ Randomized controlled trials (RCTs) are a gold standard design for internal validity in evidence-based

interventional research. This is because, in principle, at the start of the study the study groups differ in only one variable: the intervention of interest. However, there are many reasons why an RCT may have compromised internal and external validity. Some of these reasons are widely acknowledged; others, not as much so. This article observes commonly acknowledged limitations of RCTs and presents, with details and examples, less

well-recognized limitations. Earlier articles in this column are referenced, where relevant, to avoid repetition and facilitate understanding.

Commonly Acknowledged Limitations

For ethical, medicolegal, and other reasons, and to increase sample homogeneity and hence improve signal detection,² RCTs usually have narrow sample selection criteria. Thus, patients may be selected only if they meet specifications for symptoms and illness severity; they may be excluded if they are suicidal, if they have psychotic symptoms, if they have major medical or neuropsychiatric comorbidity, if they have concurrent alcohol or substance use disorder, if they also have a personality disorder, and so on. Consequently, samples in RCTs tend to be unrepresentative of patients at large. External validity in such RCTs is thereby reduced. RCTs are also commonly performed on

¹Dept. of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India.

HOW TO CITE THIS ARTICLE: Andrade C. Poorly Recognized and Uncommonly Acknowledged Limitations of Randomized Controlled Trials. *Indian J Psychol Med.* 2024;XX:1–3.

Address for correspondence: Chittaranjan Andrade, Dept. of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka 560029, India.
E-mail: andradec@gmail.com

Submitted: 14 Oct. 2024
Accepted: 22 Oct. 2024
Published Online: xxxx

 Sage



Copyright © The Author(s) 2024

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution- NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-Commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the Sage and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

ACCESS THIS ARTICLE ONLINE
Website: journals.sagepub.com/home/szj
DOI: 10.1177/02537176241297953

convenience and purposive samples and sometimes on enriched samples, further limiting their external validity.³

The internal validity of RCTs can be compromised by faulty randomization methods, poor blinding, use of assessments with uncertain reliability and validity, selection of inappropriate rating instruments, inadequate rater training, lack of standardization of methods within and across centers, and others. Problems such as inadequate recruitment and underpowering, poor treatment adherence, high dropout, high placebo response, and ceiling and floor effects can scuttle RCTs. Finally, the trustworthiness of RCT literature is compromised by poor or unscrupulous statistical methods, selective publication, selective reporting, investigator and industry bias in these regards, and others. Some of these limitations are avoidable, others are out of the control of the researcher. These limitations are well-recognized, have been discussed elsewhere, and are not repeated here.^{4,5}

Unobvious Examples of Poor External and Internal Validity

Reasons for lack of validity often hide in plain sight and are seldom discussed because attention is focused on the findings of the study and not on the bias in the sample. Consider the CONSORT diagram, scrutiny of which can often provide an idea of how filtered the sample was. In some RCTs, the filtration is so extensive that external validity is seriously compromised.^{6,8} Regrettably, the CONSORT diagram presented by authors often does not provide information about how many patients were obviously ineligible and hence not formally screened. As an example, patients who are out of age range, who are already on treatment, and who have longstanding illnesses may be obviously ineligible and not screened. In such studies, therefore, only obviously eligible subjects are screened, and despite stringent study selection criteria, most of these patients may be found eligible to enter the randomization phase. This gives the reader a false impression that few patients were found ineligible during screening and creates a false impression of external validity.⁹ Judgment about external validity should then be made

after examining the study selection criteria and not based on what the CONSORT diagram states.

Consider RCTs of interventions such as psychotherapy, yoga, meditation, acupuncture, and aerobic exercise. Patients cannot be blinded to these interventions. Patients who volunteer and/or consent for such RCTs would, therefore, experience a placebo response that is contaminated by their expectations and beliefs. As an example, patients who believe that meditation is helpful could select themselves into an RCT on meditation and experience an enhanced placebo response if they are in the meditation group or a diminished placebo response if they are in the control group. Likewise, in psychotherapy RCTs, patients in waitlisted groups would not experience a placebo response. In such RCTs, patient cooperation with study protocols and dropout could also be shaped by preexisting beliefs. Internal validity is thus seriously compromised in such RCTs.^{10,11} The subject is important in Eastern countries where interventions such as yoga and tai chi are increasingly being investigated.

In a special example of bias in RCTs of interventions that cannot be blinded, intravenous ketamine was found to be non-inferior to electroconvulsive therapy in patients with depression.¹² In this RCT, out of 403 patients who had been randomized, 31 patients dropped out in the ECT arm before starting treatment; this number was just 4 in the ketamine arm. The implication is that these dropouts were due to dissatisfaction with the assigned treatment. Thus, randomization was compromised even before the RCT began, and there is no assurance that the patients who remained in the study were free of the biases that the dropouts appeared to display.

As a less common but nonetheless influential example, RCTs have seriously compromised internal validity in maintenance therapy studies when clinically stabilized subjects are randomized to continue on active intervention or to rapidly or even abruptly switch to placebo; this has been done, for example, in studies of quetiapine, lamotrigine, and iloperidone.^{6,13,14} The compromise to internal validity is serious because an assumption of RCTs is that, at the start of the study, subjects are similar across

groups in all regards except for the allocated intervention. The assumption is violated in patients who abruptly switch to placebo because these patients experience physiological perturbations related to sudden discontinuation of active treatment, whereas patients who continue on active treatment do not experience such perturbations. The efficacy of even lithium was “established” through historical RCTs that effected such sudden switches; as we know today, the sudden discontinuation of lithium is associated with a heightened risk of early relapse.^{15,16}

Post-randomization Bias

Randomization begins to lose its integrity soon after the commencement of the trial. Most people are aware of the issues, but few realize the extent to which these issues disturb the assumption that, in RCTs, the groups are similar in all regards except for the intervention being studied. In a nutshell, the issues are events and exposures, called post-randomization biases, that disturb the balance between groups; their occurrence is also known as post-randomization confounding.¹⁷

Examples of post-randomization biases include dropout due to adverse events (more likely in the drug group), dropout due to inefficacy (more likely in the placebo group), use of rescue medication (more likely in the placebo group), and use of unreported out of trial medication (more likely in the placebo group). There is also no reason to expect that drop out due to other reasons, nonadherence to study medication, adherence to study protocols, exposure to external destabilizing events such as stress, and other biases will be neatly balanced between the groups in the study.

Some of the biases, such as those related to missing data, can be partially addressed using methods of imputation. However, to identify and address all post-randomization biases is next to impossible.¹⁷⁻¹⁹ We can recognize and accept that the RCT has been compromised to the extent that the biases were detected and reported, as with the use of rescue medication, study dropout, and reasons for study dropout. We are unable to recognize the compromised internal validity with regard to biases that are not recognized and/or not reported, as with the use of out-of-study medication and exposure to external destabilizing events.

As an example of post-randomization bias, in an RCT of vitamin D supplementation in depressed patients with vitamin D deficiency, some patients in the trial did not take the study medication, and others self-medicated with vitamin D purchased over the counter. This was detected when blood levels were checked at the end of the study. Extensive statistical reworking was required to address the problem.²⁰ One might imagine similar problems biasing results in other supplementation RCTs, such as RCTs on omega-3 fatty acids. Control group contamination may likewise occur in RCTs of educational interventions, given the ubiquitous availability of information today.

Parting Notes

When an RCT ends, treatment responders in each group may be followed for a further period²¹; such a follow-up study is a nonrandomized observational study and not the continuation phase of an RCT because selectively following responders alters the composition of the groups, thereby breaking the randomization.

The longer the duration of an RCT, the greater the likelihood of contamination by post-randomization biases. The resultant statistical noise may reduce the measured effect size of the intervention and may make a true intervention effect difficult to detect. This was apparent in a 7-year RCT of a cash intervention to reduce memory decline and dementia in an impoverished setting.²²

When RCTs measure proxies, the proxies (e.g., neuropsychological test findings) should not be misunderstood to represent clinical outcomes (e.g., cognitive symptoms in depression that impair workplace efficiency). Authors may fail to draw the distinction.²³

As final notes, limitations described in this article notwithstanding, RCTs are still the best among research designs; other research designs have even greater imperfections. Because different RCTs are different, it is important to recognize the imperfections present in each RCT so that the findings of the RCT can be soberly interpreted.

Acknowledgements

I acknowledge useful comments and suggestions on an early draft of this paper received from Dr Vikas Menon, Professor of Psychiatry,

JIPMER, Puducherry, India, and Dr Shahul Ameen, Consultant Psychiatrist, St. Thomas Hospital, Changanacherry, Kerala, India.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Declaration Regarding the Use of Generative AI

None used.

Funding

The author received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Chittaranjan Andrade  <https://orcid.org/0000-0003-1526-567X>

References

- Andrade C. Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian J Psychol Med* 2018; 40(5): 498–499.
- Andrade C. Understanding statistical noise in research: 2. Noise in clinical trials and observational studies. *Indian J Psychol Med* 2023; 45(2): 198–200.
- Andrade C. The inconvenient truth about convenience and purposive samples. *Indian J Psychol Med* 2021; 43(1): 86–88.
- Kostis JB and Dobrzynski JM. Limitations of randomized clinical trials. *Am J Cardiol* 2020; 129: 109–115.
- Mulder R, Singh AB, Hamilton A, et al. The limitations of using randomized controlled trials as a basis for developing treatment guidelines. *Evid Based Ment Health* 2018; 21(1): 4–6.
- Suppes T, Vieta E, Liu S, et al. Trial 127 Investigators. Maintenance treatment for patients with bipolar I disorder: results from a North American study of quetiapine in combination with lithium or divalproex (trial 127). *Am J Psychiatry* 2009; 166(4): 476–488.
- Kuyken W, Hayes R, Barrett B, et al. Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): a randomized controlled trial. *Lancet* 2015; 386(9988): 63–73.
- Andrade C. Examination of participant flow in the CONSORT diagram can improve the understanding of the generalizability of study results. *J Clin Psychiatry* 2015; 76(11): e1469–e1471.
- Kaur R, Sidana A, Malhotra N, et al. Oral versus long-acting injectable antipsychotic in first-episode schizophrenia: a 12 weeks interventional study. *Indian J Psychiatry* 2023; 65(4): 404–411.

- Paul-Labrador M, Polk D, Dwyer JH, et al. Effects of a randomized controlled trial of transcendental meditation on components of the metabolic syndrome in subjects with coronary heart disease. *Arch Intern Med* 2006; 166(11): 1218–1224.
- Andrade C. Transcendental meditation and components of the metabolic syndrome: methodological issues. *Arch Intern Med* 2006; 166(22): 2553.
- Anand A, Mathew SJ, Sanacora G, et al. Ketamine versus ECT for nonpsychotic treatment-resistant major depression. *N Engl J Med* 2023; 388(25): 2315–2325.
- Calabrese JR, Suppes T, Bowden CL, et al. A double-blind, placebo-controlled, prophylaxis study of lamotrigine in rapid-cycling bipolar disorder. Lamictal 614 Study Group. *J Clin Psychiatry* 2000; 61(11): 841–850.
- Weiden PJ, Manning R, Wolfgang CD, et al. A randomized trial of iloperidone for prevention of relapse in schizophrenia: the REPRIEVE study. *CNS Drugs* 2016; 30(8): 735–747.
- Verdoux H and Bourgeois M. Conséquences à court terme de l'interruption d'un traitement par sels de lithium [Short-term sequelae of lithium discontinuation]. *Encephale* 1993; 19(6): 645–650.
- Baldessarini RJ, Tondo L, Floris G, et al. Reduced morbidity after gradual discontinuation of lithium treatment for bipolar I and II disorders: a replication study. *Am J Psychiatry* 1997; 154(4): 551–553.
- Manson JE, Shufelt CL and Robins JM. The potential for post-randomization confounding in randomized clinical trials. *JAMA* 2016; 315(21): 2273–2274.
- Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can J Psychiatry* 2002; 47(1): 68–75.
- Adler AI and Latimer NR. Adjusting for nonadherence or stopping treatments in randomized clinical trials. *JAMA* 2021; 325(20): 2110–2111.
- Kumar PNS, Menon V and Andrade C. A randomized, double-blind, placebo-controlled, 12-week trial of vitamin D augmentation in major depressive disorder associated with vitamin D deficiency. *J Affect Disord* 2022; 314: 143–149.
- Jelovac A, Kolshus E and McLoughlin DM. Relapse following bitemporal and high-dose right unilateral electroconvulsive therapy for major depression. *Acta Psychiatr Scand* 2021; 144(3): 218–229.
- Rosenberg M, Beidelman ET, Chen X, et al. Effect of a cash transfer intervention on memory decline and dementia probability in older adults in rural South Africa. *Proc Natl Acad Sci USA* 2024; 121(40): e2321078121.
- Manna CK, Ranjan R, Kumar P, et al. Effect of vortioxetine versus venlafaxine on cognitive functions in adults with major depressive disorder: a randomized controlled trial. *Indian J Psychiatry* 2023; 65(8): 815–824.