

# Positive thinking about negative studies

Eva Petkova, Adam Ciarleglio, Patricia Casey, Norman Poole, Kenneth Kaufman, Stephen M. Lawrie, Gin Malhi, Najma Siddiqi, Kamaldeep Bhui and William Lee

## Summary

The non-reporting of negative studies results in a scientific record that is incomplete, one-sided and misleading. The consequences of this range from inappropriate initiation of further studies that might put participants at unnecessary risk to treatment guidelines that may be in error, thus compromising day-to-day clinical practice.

## Keywords

Adjustment disorders; aetiology; alcohol disorders; anorexia nervosa; anthropology.

## Copyright and usage

© The Author(s), 2024. Published by Cambridge University Press on behalf of Royal College of Psychiatrists.

Colloquially, studies that do not provide sufficient statistical evidence in support of what the investigators hypothesised are termed ‘negative studies’. A great deal can be learned from negative studies, but such studies are less frequently published. For example, about half of the conducted clinical studies remain unpublished<sup>1</sup> and negative studies are more likely to be unpublished than studies that provided sufficient evidence in support of the research hypothesis.<sup>2</sup> See Chalmers & Glasziou<sup>3</sup> for an overview.

## What are negative studies and why are they not published?

Many research questions can be answered by formally setting up an alternative hypothesis (what we aim to verify) against a null hypothesis (the status quo) and evaluating the evidence against the null. The logic of this framework is that what we are testing would be considered a viable reality only if there is strong evidence against the status quo (the null hypothesis) and in favour of this alternative. The collected data will either be consistent with the null hypothesis or they will not. If the data are consistent with the null, it is said that the study ‘failed to reject the null hypothesis’, i.e. the status quo remains; if the data are not consistent with the null hypothesis but are supportive of the alternative, then it is said that the study ‘rejected the null hypothesis’ in favour of the alternative, i.e. the status quo is rejected in favour of the hypothesised alternative. The former is referred to as a ‘negative’ study and the latter as a ‘positive’ study. This binary presentation is subject to uncertainty, often quantified by the probability for type I error (to falsely reject the null hypothesis when it is true) and type II error (to falsely fail to reject the null when the alternative hypothesis is true) of the tests employed.

We propose that the main reasons for negative studies remaining unpublished are twofold. First, journal editors and reviewers may favour positive studies because they appear to be presenting new knowledge by challenging the status quo and thus are potentially of greater interest to readers and may therefore be cited more frequently. Second, and probably more importantly, investigators and their sponsors may feel their time and effort are not well spent on drafting and submitting papers for publication that are more likely to be rejected or remain relatively uncited in comparison with other work.<sup>4</sup>

## Why publish negative studies?

Non-reporting of negative studies results in a published record that is incomplete, one-sided and misleading. One consequence of this publication bias is that it may inappropriately necessitate further studies, putting participants of these studies at unnecessary risk and also

causing avoidable burden to researchers and institutions. This is unethical and could undermine the trust of those participating in clinical research – a trust that is vital for its continuation and growth. Misleading literature also means that treatment guidelines, and therefore day-to-day clinical practice, may be in error. People may receive unnecessary treatments and suffer negative outcomes as a consequence.

Results from negative studies are therefore an essential piece of the totality of scientific evidence. Consider four scenarios under which a negative study may arise:

- the study was well-designed, adequately powered, faithfully completed and correctly analysed
- the study was well-designed, adequately powered at the design phase and correctly analysed, but did not achieve the planned sample size and was therefore underpowered to detect the anticipated effect
- the study was well-designed but inadequately powered at the design phase
- the study was neither well-designed nor adequately powered.

With the possible exception of case (d), there is value in having the results from all the above scenarios available to the scientific community. Under scenario (a), a negative result could mean that the true effect is smaller than the investigator had thought to be clinically meaningful and had powered the study to detect, or, if the study was planned as a replication, that the previous results were not replicated. Alternatively, it could mean that the true effect is of the size anticipated by the investigators, but that type II error occurred. This information, along with information from other studies of the same relationship, can be combined in a meta-analysis to provide a pooled estimate of the effect from the available scientific evidence. The information could also be used to plan future studies of the effect of interest or help investigators to make well-informed decisions regarding directions for their own research. In such cases it is essential that researchers conduct mechanistic and process evaluation to learn more about the reasons for the negative findings.

Scenario (b) is similar to (a), except now the chance for type II error is larger than what the investigators intended since the study did not achieve the planned sample size. It is also possible that the true effect is smaller than the investigators expected, or, in case of a replication, that the previously reported results were not replicated, but because the study was underpowered, the confidence interval around the estimate of the effect is larger than planned and the interpretation of the results is more difficult.

Scenario (c) describes studies that could discover only large effects in situations where smaller effects are of clinical importance. Reporting such studies is less informative than the studies in scenarios (a) and (b) because, in addition to potential type II

error, it only tells us that the effect is not as large as the investigators expected, but that does not necessarily imply the absence of clinically meaningful effect. Reporting the study indicates to the research community that it was conducted, and the estimated effect of interest, with its respective confidence interval, can be used to inform the sample size calculations for future research, akin to estimates that are derived from pilot studies. Scenarios (a), (b) and (c) could contribute to meta-analyses that might allow for research questions to be adequately answered even when none of the studies alone were able to do this. Scenario (d) describes studies that probably do not provide any value, unless the reports critique in detail all aspects of the study design and conduct, which might help future investigators avoid the same potential design and execution pitfalls.

### Powering of studies: consequences of type I and II errors

Medical studies are typically planned with smaller risk of type I errors (commonly 5%) than that of type II errors (commonly 20%). This practice is consistent with the view that the consequences of declaring the existence of an effect when an effect does not exist are worse than the consequences of failing to detect an effect and preserving the status quo when an effect exists.

Type I error might mean mistakenly going against the current practice and, for example, adopting new treatments or policies that are ineffective, wasteful or harmful. Type II error might mean failing to correctly reject the current practice, which can result in missed opportunities for innovation and can also have dire practical consequences. Accordingly, we advocate for careful consideration of how studies should be powered, depending on the context, with particular consideration of the real-world consequences of type II errors.

Inadequate sample size is an important factor in mental health research. Unlike the areas of cancer and cardiovascular disease, mental health (as well as other specialties) lacks the international multisite collaborations that would facilitate large trials and deep phenotyping involving tens of thousands of participants. Although there are some exceptions, such as data repositories for images from resting-state functional magnetic resonance imaging,<sup>5</sup> standardised collections of images under specific neuropsychiatric tasks contain data on only a limited number of patients. In the same way, central nervous system RNA and DNA studies on post-mortem brains of patients with mental health conditions are limited by the small number of samples.<sup>6</sup> Large-scale definitive studies and trials in other specialties have yielded good results and we look forward to similar steps in mental health research.

Reporting negative studies should follow the standards set by the CONSORT Group<sup>7</sup> (also see <https://www.equator-network.org/reporting-guidelines/consort/>). Specific emphasis should be put on discussing: (a) the target effect size that was used for determining the sample size for the study, together with the rationale for selecting that effect size; (b) what were the acceptable type I and type II errors, with justification; (c) whether the planned sample size was achieved; (d) how the results should be interpreted, including in the context of existing evidence, if relevant.

### Solutions and challenges







What can be done to ensure that negative results are available to the scientific community? First, journal editors need to give notice to reviewers and associate editors that the journal is dedicated to publishing manuscripts describing impactful and timely well-designed and well-executed studies, whether the result is positive or negative, based on conventional statistical criteria. Second,

journal editors need to convey this message to authors by including it in the journals' instructions for authors. Third, journals should highlight these studies when they are published to show that there is value in the honest reporting of well-designed and executed studies regardless of their results. Fourth, authors need to pursue writing up and submitting these manuscripts to journals. Fifth, investigators need to pay careful attention to citing negative studies to remove some of the forces that create and maintain this problem.<sup>8</sup>

Many initiatives now exist to improve the probability of publication of negative studies. One such is 'Registered Reports', wherein the Introduction and Methods sections of papers are reviewed by a journal before the data are collected.<sup>9</sup> This allows independent reviewers to have influence at the most effective point in the process, before the study is undertaken. After revisions, if those sections are accepted, the final paper is guaranteed publication once the study is complete. This makes it impossible for the journal or the researcher to be influenced in their decisions by the results of the study. One journal implementing this process is the *Journal of the American Academy of Child and Adolescent Psychiatry*; see also [www.cos.io/initiatives/registered-reports](http://www.cos.io/initiatives/registered-reports). Although not the current editorial policy of the BJPsych portfolio of journals, Registered Reports might be a potential future direction. This approach brings some benefits, but may be criticised because it may consume more of the limited, and usually unpaid, time of journal editors and reviewers. Further, the Registered Reports approach does not acknowledge the qualitative and quantitative properties of future scientific results, which may affect the potential impact of a paper. This then may affect whether a widely read journal (itself competing for citations under current paradigms) will be likely to accept it.

Further challenges to the current independent journal-based model of publication of results of trials exist. Funders themselves often publish short reports on their own websites; authors use preprint servers; one funder, the Wellcome Trust, has recently undertaken to guarantee publication of research it funds. These efforts are a worthwhile challenge to scientific publishers, and may yet bring an end to what we recognise as a learned journal. However, independent, often anonymous, peer-review remains the gold standard of quality in scientific publishing and provides some assurance to readers. For high standards, it is important that no one 'marks their own homework'.

Ultimately, the validity of the scientific enterprise is in the hands of all of us, editors, reviewers, authors and funders, who control what is submitted and published and what remains unseen. We must recognise our roles in these decisions and need to promote a culture that puts high value on encouraging and upholding the completeness of scientific knowledge.

**Eva Petkova**, NYU Grossman School of Medicine, New York University, New York, USA; **Adam Ciarleglio**, George Washington University School of Public Health and Health Services, Washington, DC, USA; **Patricia Casey** , Hermitage Medical Clinic, Dublin, Ireland; and Department of Psychiatry, University College Dublin, Dublin, Ireland; **Norman Poole** , Department of Neuropsychiatry, South West London and St George's Mental Health NHS Trust, London, UK; **Kenneth Kaufman** , Department of Psychiatry, Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA; and Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; **Stephen M. Lawrie** , Department of Psychiatry, University of Edinburgh, Edinburgh, UK; **Gin Malhi** , Academic Department of Psychiatry, Kolling Institute, Northern Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia; CADE Clinic and Mood-T, Royal North Shore Hospital, Northern Sydney Local Health District, Sydney, New South Wales, Australia; and Department of Psychiatry, University of Oxford, Oxford, UK; **Najma Siddiqi** , Department of Health Sciences, Hull York Medical School, University of York, York, UK; **Kamaldeep Bhui** , Department of Psychiatry, Nuffield Department of Primary Care Health Science, University of Oxford, Oxford, UK; Wadham College, University of Oxford, Oxford, UK; East London and Oxford Health NHS Foundation Trusts, London, UK; and WPA Collaborating Centre Oxford, Oxford, UK; **William Lee** , Department of Liaison Psychiatry, Cornwall Partnership NHS Trust, Bodmin, UK

**Correspondence:** Eva Petkova. Email: [eva.petkova@nyulangone.org](mailto:eva.petkova@nyulangone.org)

First received 14 Apr 2023, final revision 18 Oct 2023, accepted 25 Oct 2023

### Data availability

Data availability is not applicable to this article as no new data were created or analysed in this study.

### Author contributions

All authors contributed to the preparation and further development of this article, both before and after peer review, and all have seen and approved the final version.

### Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

### Declaration of interest

E.P. is a Deputy Editor of *BJPsych Open* and Statistical Advisor for RCPsych journals; P.C. is Editor in Chief of *BJPsych Advances*; K.K. is Editor in Chief of *BJPsych Open*; S.M.L. and N.S. are members of the *BJPsych* editorial board; G.M. is Editor in Chief of the *BJPsych* and a member of the *BJPsych Open* editorial board; K.B. is a member of the *BJPsych* editorial board; W.L. is a Deputy Editor of the *BJPsych*; all authors except for A.C. are members of the Royal College of Psychiatrists' Research Integrity Group, of which W.L. is Chair. None of the authors took part the review or decision-making process of this paper.

### References

- 1 Scherer RW, Meerpohl JJ, Pfeifer N, Schmucker C, Schwarzer G, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2018; **11** (<https://doi.org/10.1002/14651858.MR000005.pub4>).
- 2 Kicinski M, Springate DA, Kontopantelis E. Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. *Stat Med* 2015; **34**: 2781–93.
- 3 Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Obst Gynecol* 2009; **114**: 1341–5.
- 4 Hopewell S, Dickersin K, Clarke MJ, Oxman AD, Loudon K. Publication bias in clinical trials. *Cochrane Database Syst Rev* 2007; **2**: MR000006 (<https://doi.org/10.1002/14651858.mr000006.pub2>).
- 5 Nooner KB, Colcombe SJ, Tobe RH, Mennes M, Benedict MM, Moreno AL, et al. The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Front Neurosci* 2012; **6**: 152.
- 6 McCullumsmith R, Hammond J, Shan D, Meador-Woodruff JH. Postmortem brain: an underutilized substrate for studying severe mental illness. *Neuropsychopharmacol* 2014; **39**: 65–87.
- 7 Butcher NJ, Monsour A, Mew EJ, Chan AW, Moher D, Mayo-Wilson E, et al. Guidelines for reporting outcomes in trial reports: the CONSORT-Outcomes 2022 extension. *JAMA* 2022; **328**: 2252–64.
- 8 Greenberg S. How citation distortions create unfounded authority: analysis of a citation network. *BMJ* 2009; **339**: b2680.
- 9 Chambers CD, Tzavella L. The past, present and future of Registered Reports. *Nat Hum Behav* 2022; **6**: 29–42.