EDITORIAL

When beauty is but skin deep: dealing with problematic studies in systematic reviews

Stephanie L Boughton, Jack Wilkinson, Lisa Bero

Cochrane Database of Systematic Reviews 2021;(6):ED000152 **https://doi.org/10.1002/14651858.ED000152**

Publication date: 3 June 2021

Systematic reviews provide high-quality assessment of the evidence for healthcare interventions, but this assessment is predicated on the trustworthiness of the studies they include. The presence of untrustworthy, or 'problematic', studies in the literature poses challenges for systematic reviewers and is a threat to the reliability of systematic reviews. This is compounded by the absence of an agreed definition of what constitutes a problematic study and the lack of a validated method to identify them.

The most visible problematic studies are retracted studies. The number of retractions has risen rapidly,[1][2][3] suggesting a rise in problematic studies in the literature or an increase in awareness of them. Articles may be retracted for a number of different reasons, including error, naïve mistakes, or research misconduct,[4] and estimates vary as to the percentage of retractions that are due to misconduct versus honest error. One study found 67.4% of retractions were due to misconduct (comprised of fraud, suspected fraud, duplicate publication, and plagiarism).[5]

The MECIR standards (Methodological Expectations of Cochrane Intervention Reviews; community.cochrane.org/mecir-manual) are clear that retracted articles should not be included in systematic reviews, but retracted studies are likely to represent only the tip of the iceberg. A systematic review found that 2% of scientists admitted having fabricated, falsified, or modified data or results, while 14% reported having witnessed a colleague doing so,[6] suggesting the scale of problematic studies is far greater than the number of articles retracted. In an assessment by John Carlisle of 526 randomized controlled trials submitted to the journal *Anaesthesia* between February 2017 and March 2020, he believed 8% to be fatally flawed due to containing false data.[7] For those where he had access to individual patient data, this rose to 44% containing false data, and 26% being classified as critically flawed. It is unclear how many problematic studies go undetected during typical journal editorial processes, which do not usually include the sort of forensic investigation undertaken by Carlisle.

Even studies that are eventually retracted have downstream effects on the healthcare literature, including on systematic reviews, due to the time lag between their initial publication and subsequent retraction. A study of all retractions indexed in PubMed up to May 2012 found that the average time-to-retraction was 32.91 months, with a range of less than 1 month to 304 months.[8] Another study of retracted obstetrics and gynaecology articles up to June 2018 found a median time to retraction of two years, with one study being retracted 18 years after publication.[3] During this time there may be no indication to readers or those conducting a systematic review for which it is eligible for inclusion that there are any concerns about the study, though use of Expressions of Concern to alert readers to potential issues is increasing.[9]

Cochrane has developed a new policy to address potentially problematic studies in the context of systematic reviews. The Cochrane policy for managing potentially problematic studies and accompanying implementation guidance set out how Cochrane Review authors and Cochrane Review Groups (CRGs) should manage situations involving potentially untrustworthy, or 'problematic' studies in the context of Cochrane Reviews. It provides steps to follow when an included study is retracted or has a formal Expression of Concern published, and also how to deal with scenarios where there is no formal post-publication amendment but review authors or the CRG have concerns about the trustworthiness of the data in an included study, or study eligible for inclusion in a review.

While this new policy provides guidance for review authors and editors where they have identified serious concerns about a study, there is an urgent need for a validated method to identify problematic studies reliably and fairly. Existing tools for assessing included studies, such as tools for assessing risk of bias, assume that the data are real. A variety of methods to identify problematic studies have been proposed. One example is the REAPPRAISED checklist for evaluation of publication integrity.[10] Suggested methods range from testing for statistical anomalies, searching for inconsistencies in how data are reported (for example different sample sizes for the same study reported in a publication, or across several publications), searching for awkward or repetitive phrasing, analyzing images, and scrutinizing all publications from authors who have previously published retracted or problematic studies. However, many of these methods involve subjective evaluation of data, and none have been fully evaluated. Use of unvalidated methods risks over- or under-detection of problematic studies, and caution is needed as misclassification of a genuine study as problematic could result in erroneous review conclusions. Misclassification could also lead to reputational damage to authors, legal consequences, and ethical issues associated with participants having taken part in research, only for it to be discounted.

The scope of problematic studies encompasses a wide range of different issues, varying in scale and cause. They may involve individual articles, groups of articles by a particular author or group of authors, or systematic manipulation of the publication process where individuals or groups use fraudulent or dishonest practices to inappropriately influence the publication process,[11] for example so called 'paper mills', where authors allegedly buy services such as ghostwritten fraudulent or fabricated manuscripts produced by a third party.[12] Consensus is greatly needed on terminology and what constitutes a 'problematic' study. For the purposes of the Cochrane policy, we have defined a problematic study as "any published or unpublished study where there are serious questions about the trustworthiness of the data or findings, regardless of whether the study has been formally retracted", but we know that such terms mean different things to different individuals, denoting a greater or lesser degrees of 'seriousness'. As well as having different thresholds for what they consider concerning, individuals will also have different 'red flags' for what they believe is indicative of untrustworthiness.

The challenge ahead is to establish consensus around reliable, feasible approaches for identifying whether a study is problematic in the first place. As a starting point we need a standard, agreed definition of a 'problematic' study. Research is then needed to develop and validate methods to identify such problematic studies, in order to preserve the status of systematic reviews as a source of definitive health information.

Correction (9 July 2021): The third paragraph of this article has been amended to state that John Carlisle assessed 526 randomized controlled trials submitted to the journal Anaesthesia, rather than Anaesthesiology as was previously incorrectly stated.

Author Information

Stephanie L Boughton[1], Jack Wilkinson[2], Lisa Bero[3]

[1]Editorial and Methods Department, Cochrane, UK. [2]Centre for Biostatistics, Manchester Academic Health Science Centre, Faculty of Biology, Medicine, and Health, University of Manchester, UK. [3]Center for Bioethics and Humanities, University of Colorado Anschutz Medical Campus, USA

Declarations of interest

SLB is Research Integrity Editor at Cochrane and was involved in the development of the Cochrane policy for managing potentially problematic studies.

JW is a Statistical Editor for Cochrane Gynaecology and Fertility, and declares that publications of this nature could plausibly benefit his career.

LB is Senior Editor, Research Integrity, Cochrane and was involved in the development of the Cochrane policy for managing potentially problematic studies.

Provenance and peer review

This editorial was commissioned and was not externally peer reviewed.

References

1. Wager E, Williams P. Why and how do journals retract articles? An analysis of Medline retractions 1988–2008. *Journal of Medical Ethics* 2011;37:567–70. http://dx.doi.org/10.1136/jme.2010.040964

2. Steen RG. Retractions in the scientific literature: is the incidence of research fraud increasing? *Journal of Medical Ethics* 2011;37:249–53. https://doi.org/10.1136/jme.2010.040923

3. Chambers LM, Michener CM, Falcone T. Plagiarism and data falsification are the most common reasons for retracted publications in obstetrics and gynaecology. *BJOG* 2019;126:1134–40. https://doi.org/10.1111/1471-0528.15689

4. COPE Council. COPE retraction guidelines. November 2019. https://doi.org/10.24318/cope.2019.1.4

5. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences USA* 2012;109(42):17028–33. https://doi.org/10.1073/pnas.1212247109

6. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS One* 2009;4(5):e5738. https://doi.org/10.1371/journal.pone.0005738

7. Carlisle JB. False individual patient data and zombie randomised controlled trials submitted to Anaesthesia. *Anaesthesia* 2020;76(4):472–9. https://doi.org/10.1111/anae.15263

8. Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased? *PLOS One* 2013;8(7):e68397. https://doi.org/10.1371/journal.pone.0068397

9. Vaught M, Jordan DC, Bastian H. Concern noted: a descriptive study of editorial expressions of concern in PubMed and PubMed Central. *Research Integrity and Peer Review* 2017;2:10. https://doi.org/10.1186/s41073-017-0030-2

10. Grey A, Bolland MJ, Avenell A, Klein AA, Gunsalus CK. Check for publication integrity before misconduct. *Nature* 2020;577:167–9. https://doi.org/10.1038/d41586-019-03959-6

11. COPE Council. Systematic manipulation of the publication process. Version 1. 2018. https://doi.org/10.24318/cope.2019.2.23

12. Byrne J, Christopher J. Digital magic, or the dark arts of the 21st century—how can journals and peer reviewers detect manuscripts and publications from paper mills? *FEBS Letters* 2020;594(4):583–9. https://doi.org/10.1002/1873-3468.13747