Publication Bias: The Elephant in the Review

Jarrod E. Dalton, PhD,*† Shari D. Bolen, MD, MPH,‡ and Edward J. Mascha, PhD*†

Systematic reviews and meta-analyses serve an important role in summarizing the results of multiple investigations, particularly rigorously designed trials aimed at the estimation of specific treatment effects. It is widely believed that, when properly conducted, results of meta-analyses can be stronger than single investigations because of increased sample sizes and diversity of settings. Proper conduct of these investigations, however, is simultaneously difficult and labor intensive. Clearly, the conclusion reached by a particular meta-analysis or systematic review depends on many methodologic factors, including choice of search criteria, choice of literature sources, study inclusion and exclusion criteria, procedures for abstracting treatment effects, and, importantly, the extent to which things like confounding might differentially affect the results from included studies.

One key factor that may affect the conclusions reached by many such reviews is the hidden elephant of publication bias. In this issue of Anesthesia & Analgesia, Hedin et al1 provide an assessment of the extent to which systematic reviews and meta-analyses included evaluations of publication bias. Describing publication bias as the tendency to publish “only results that are statistically or clinically significant,” they found that, among 207 systematic reviews meeting inclusion criteria, only 114 (55%) discussed it and 89 (43%) evaluated it. Furthermore, they found that only 68 (33%) of the reviews reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (a number that may be artificially low as only 68 (33%) of the included reviews reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines). The distinction is important when evaluating publication bias because systematic reviews by definition do not typically contain an assessment of overall treatment effect. We cannot verify the authors’ claims that the proportion of anesthesia reviews evaluating publication bias may affect results is lower among this subset. Indeed, based on our review of Table 4, it appears to us that a rather small number (say, 2–5) of the included reviews had meaningful differences in pooled effect estimates after accounting for publication bias. Regardless of this point, however, publication bias does exist and can lead to large differences in conclusions when not adequately assessed and addressed.

Hedin et al briefly distinguished systematic review and meta-analyses in the title and in the Methods section but subsequently referred to all articles as “systematic reviews” in the Results and Discussion sections. Systematic reviews are a reproducible method for identifying and collating all empirical evidence on a specific topic. These may or may not include meta-analyses, which is a quantitative method for combining individual studies to estimate an overall treatment effect. The distinction is important when evaluating publication bias because systematic reviews by definition do not typically contain an assessment of overall treatment effect. We cannot verify the authors’ claims that the proportion of anesthesia reviews evaluating publication bias is lower than that of other medical fields or that the prevalence of publication bias is greater than that of other medical fields, because their estimates are not restricted to reviews that included meta-analyses.

Appreciating the mechanisms by which publication bias and other forms of reporting bias might arise is a prerequisite to minimizing their ultimate impact on our interpretation of the literature and correspondingly needs to be thoroughly understood by the anesthesia research community. Such mechanisms include confirmation bias (selective preference for new results that agree with previous evidence), improper study design (eg, lack of power for detecting meaningful differences), improper specification of the relevant patient population for the intervention of interest), improper hypothesis

From the Departments of *Quantitative Health Sciences and †Outcomes Research, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, Ohio; and ‡Division of Internal Medicine, Center for Health Care Research and Policy, MetroHealth Medical Center/Case Western Reserve University, Cleveland, Ohio.

Accepted for publication July 26, 2016.

Funding: This study was provided, in part, by the Clinical and Translational Science Collaborative of Cleveland, KL2TR000440 from the National Center for Advancing Translational Sciences component of the National Institutes of Health (NIH) and the NIH roadmap for Medical Research. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The funding agreement ensured the authors’ independence in designing the study, interpreting the data, writing and publishing the report.

The authors declare no conflicts of interest.

Reprints will not be available from the authors.

Address correspondence to Jarrod E. Dalton, PhD, Cleveland Clinic, 9500 Euclid Ave (JNJ3), Cleveland, OH 44195. Address e-mail to daltonj@ccf.org.

Copyright © 2016 International Anesthesia Research Society

DOI: 10.1213/ANE.0000000000001596
testing practice (discontinuation of the research/manuscript development process once negative results are established in the analytic phase), lack of appropriate avenues for reporting negative studies beyond “gray” literature (such as dedicated journal sections that report negative findings), selective outcomes reporting bias (reporting only outcomes with positive findings), selective analyses reporting bias (only reporting analyses that led to positive findings), and lack of incentives for completion of mechanistic research studies (such as requirements of advance registration by journals).

Ultimately, both investigators and editorial boards bear responsibility for ensuring proper conduct so that systematic reviews and meta-analyses achieve the ideal of providing the highest quality evidence on efficacy. As Hedin et al detail, requirements for reporting systematic reviews are inconsistent and generally scant in anesthesia journals. We agree with Hedin et al that anesthesia journals should require that authors follow the PRISMA2 (or, similarly, Meta-analysis of Observational Studies in Epidemiology5) guidelines when submitting systematic reviews, both in their instructions for authors and in implementation of the peer review process. At a minimum, authors of systematic reviews should be required to include a PRISMA checklist and flow diagram. We suggest that anesthesia journals might establish sections for negative studies, but at least dedicate sufficient space for publishing them, assuming they are of sufficient quality.

Investigators also should improve the quality of systematic reviews and meta-analyses. First, they should include a “gray” literature search to assist with reducing publication bias, although care should be taken when incorporating nonpeer-reviewed sources such as conference abstracts that can oftentimes be works in progress, pilot studies, or in the worst case, incorrect. If evaluating pharmaceuticals or medical devices, we recommend searching the Food and Drug Administration’s website in addition to clinical trials.gov and other sources mentioned by the authors.

Second, investigators should use appropriate techniques to assess publication bias, such as Egger regression or symmetry of funnel plots whenever there are >10 studies combined in a meta-analysis (with <10 the assessment methods are not very reliable). One caveat in assessing publication bias is to ensure that the nature of the observed publication bias is well understood. For instance, an asymmetric funnel plot could be the result of missing small positive studies (instead of the typical small negative studies) and would likely strengthen any positive conclusions as opposed to weakening the conclusion.

Third, investigators should measure the effect of publication bias on the estimated treatment effect(s) whenever publication bias is suspected. Particularly, we recommend the use of Duval and Tweedie’s trim-and-fill technique, which aims to adjust pooled treatment effect estimates to account for funnel plot asymmetry and which the authors helpfully demonstrate. When heterogeneity in treatment effects across studies is suspected (or indicated by the I2 statistic8), publication bias should be evaluated within meaningful a priori defined subpopulations across which treatment effects might vary.

Another technique for assessing publication bias, known as the “fail-safe N method” or “Rosenthal analysis,”7 involves identifying the number of additional negative studies that would be needed to increase the P value in a meta-analysis to above .05. The fail-safe N method is popular because it is simple to apply and purports to measure the fragility of reported findings; however, it is highly dependent on the treatment effects assumed for unobserved studies, ie, it can vary greatly based on what is assumed and is thus a fragile number itself. More importantly, it inherently places too much emphasis on the P value and its arbitrary threshold (usually P < .05) instead of focusing on the estimated treatment effects and confidence intervals, which give considerably more information. We, therefore, concur with the authors of the Cochrane handbook in recommending against the use of the fail-safe N method.8

More broadly, our research community needs to move past our collective tendency to give preferential treatment to positive results. This aggregate behavior might be because of a false sense of belief that—assuming studies are well designed and well executed—there is more to learn from positive findings than from negative findings. This may be because of the long history of positive research findings in medicine being translated to improved quality of life and longevity or due to the fact that positive studies are more likely to be published in high-impact journals2 and cited more often (ultimately resulting in increased impact factors for the journals). Regardless, we should continue to work toward creating (and acculturating to) a professional environment that focuses on the quality of research questions and the quality of research methods, independent of study findings. Assessment of publication bias should become more routine as we move forward, and finding it should become less routine.

DISCLOSURES

Name: Jarrod E. Dalton, PhD.

Contribution: This author helped write the manuscript.

Name: Shari D. Bolen, MD, MPH.

Contribution: This author helped write the manuscript.

Name: Edward J. Mascha, PhD

Contribution: This author helped write the manuscript.

This manuscript was handled by: Thomas R. Vetter, MD, MPH.

REFERENCES


