CHAPTER 1

Publication Bias in Meta-Analysis

Hannah R. Rothstein
Department of Management, Zicklin School of Business, Baruch College, New York, USA

Alexander J. Sutton
Department of Health Sciences, University of Leicester, UK

Michael Borenstein
Biostat, Inc., USA

PUBLICATION BIAS AS A THREAT TO VALIDITY

Publication bias is the term for what occurs whenever the research that appears in the published literature is systematically unrepresentative of the population of completed studies. Simply put, when the research that is readily available differs in its results from the results of *all* the research that has been done in an area, readers and reviewers of that research are in danger of drawing the wrong conclusion about what that body of research shows. In some cases this can have dramatic consequences, as when an ineffective or dangerous treatment is falsely viewed as safe and effective. This can be illustrated by two events that received much media attention as this book was going to press in late 2004. These are the debate surrounding Merck's recall of Vioxx, a popular arthritis drug (Merck maintained that it recalled Vioxx as soon as the data indicated the high prevalence of cardiovascular events among those who took Vioxx for more than 18 months, while media reports said that Merck hid adverse event evidence for years), and the use of selective serotonin reuptake inhibitor (SSRI) anti-depressants among adolescents (Elliott Spitzer, attorney general of New York State, filed a 2004 lawsuit against GlaxoSmithKline,

Publication Bias in Meta-Analysis – Prevention, Assessment and Adjustments Edited by H.R. Rothstein, A.J. Sutton and M. Borenstein

charging that they had concealed data about the lack of efficacy and about the increased likelihood of suicide associated with the use of Paxil for childhood and adolescent suicide). In most cases, the causes of publication bias will not be as clear, nor the consequences as serious as in these examples. Nevertheless these examples highlight why the topic is critically important.

Publication bias is a potential threat in all areas of research, including qualitative research, primary quantitative studies, narrative reviews, and quantitative reviews, that is, meta-analysis. Although publication bias has likely been around for as long as research has been conducted and reported, it has come to prominence in recent years largely with the introduction and widespread adoption of the use of systematic review and meta-analytic methods to summarize research. In part, this is because, as methods of reviewing have become more scientific and quantitative, the process of reviewing (and synthesizing) research has been increasingly seen as paralleling the process of primary research. Parallels to the threats to the validity of primary research have been uncovered at every step of the systematic review process (Cooper, 1998; Shadish *et al*., 2002). Furthermore, as methods of reviewing have become more systematic and quantitative, it has been possible to empirically demonstrate the existence of publication bias and to quantify its impact. Thus, a problem that was viewed hazily through the looking glass of traditional reviews came into sharp focus under the lens of meta-analysis.

In meta-analysis, publication bias is a particularly thorny issue because meta-analysis has been put forward as providing a more accurate appraisal of a research literature than is provided by traditional narrative reviews (Egger *et al*., 2000), but if the sample of studies retrieved for review is biased, then the validity of the results of a meta-analytic review, no matter how systematic and thorough in other respects, is threatened. This is not a hypothetical issue: evidence that publication bias has had an impact on meta-analyses has been firmly established by several lines of research (see Chapters 2 and 10 of this volume).

Since systematic reviews are promoted as providing a more objective appraisal of the evidence than traditional narrative reviews, and since systematic review and meta-analysis are now generally accepted in many disciplines as the preferred methodology for summarizing a literature, threats to their validity must be taken very seriously. Publication bias must be taken especially seriously, as it presents perhaps the greatest threat to the validity of this method. On the other hand, the vulnerability of systematic review and meta-analysis to publication bias is not an argument against their use, because such biases exist in the literature irrespective of whether systematic review or other methodology is used to summarize research findings. In fact, we suggest that the attention given to objectivity, transparency and reproducibility of findings in systematic reviews and meta-analyses has led to the first serious attempt to confront the problems that have always existed because of publication biases, and to ameliorate them. As demonstrated by this volume, there are now several tools available with which meta-analysts can assess the potential magnitude of bias caused by selective publication. When the potential for severe bias exists in a given analysis, this can now be identified, and appropriate cautionary statements about the meta-analytic results can be made. When potential bias can effectively be ruled out, or shown not to threaten the results and conclusions

of a meta-analysis, the validity and robustness of these results and conclusions are strengthened.

Publication bias was originally defined as the publication or non-publication of studies depending on the direction and statistical significance of the results, and the first systematic investigations of publication bias focused on this aspect of the problem. However, as readers will appreciate as they work through the book, there are numerous potential information suppression mechanisms that go well beyond the simple definition given above, including: language bias (selective inclusion of studies published in English); availability bias (selective inclusion of studies that are easily accessible to the researcher); cost bias (selective inclusion of studies that are available free or at low cost); familiarity bias (selective inclusion of studies only from one's own discipline, and outcome bias (selective reporting by the author of a primary study of some outcomes but not others, depending on the direction and statistical significance of the results). All of these biases lead to the same consequence, namely that the literature located by a systematic reviewer will be unrepresentative of the population of completed studies; hence all present the same threat to a review's validity. For this reason, it has been suggested that a single, broadly encompassing term, dissemination bias (Song *et al*., 2000), be used to refer to the problem. We agree with this sentiment, but the widespread and established use of the term publication bias has made us hesitant to tamper with, and potentially confuse, the current terminology. Readers should bear in mind that when they read 'publication bias' the broader but more cumbersome 'publication bias and associated dissemination biases' is implied.

ORGANIZATION OF THE BOOK

The book is split into three parts, and there are three appendices. Part A contains a set of chapters which together provide a non-technical introduction to publication bias and describe how it can be minimized in future research. Part B presents each of the currently available methods for assessing or adjusting for publication bias in a meta-analytic context; these chapters also illustrate each method using the data sets described in Appendix A. The chapters in Part C discuss several advanced and emerging issues that have not yet received much attention elsewhere in the literature. Finally, Appendix B is an annotated bibliography that provides illuminating further reading on publication bias; it is presented in chronological order to allow the reader to see how the field has developed over time. While we have attempted to compile and edit the book so that the chapters are integrated (with as much detailed cross-referencing of chapters as possible), we believe that each chapter can stand on its own. A more detailed outline of the contents of each section of the book is given below.

Outline of Part A

In Chapter 2, Kay Dickersin begins with a thoughtful review of the causes and origins of publication bias, after which she presents a comprehensive overview

of the history and findings of publication bias research. Her chapter concludes with some suggestions for minimizing publication bias in the future, which are elaborated upon in Chapter 3 by Jesse Berlin and Davina Ghersi. These authors propose two strategies which, if widely adopted, would go a long way towards alleviating publication bias in trials-based research areas. The first is prospective registration of clinical trials, which would create an 'unbiased sampling frame for subsequent meta-analyses'. As Berlin and Ghersi point out, however, to avoid publication bias, this would need to be coupled with open access to the results of these trials. The second is prospective meta-analysis, whereby multiple groups of investigators conducting ongoing trials agree, prior to knowing the results of their studies, to combine their findings when the trials are complete. In a variant of this strategy, the meta-analysis is *designed* prospectively to standardize the instruments used to measure specific outcomes of interest across studies. In Chapter 4, Sally Hopewell, Mike Clarke and Sue Mallett describe how to minimize publication bias by attempting to locate and retrieve grey and unpublished literature. They also point out the problems associated with the retrieval and inclusion of this literature, namely that it is time-consuming and difficult, and that its methodological quality can be hard to assess. They conclude by suggesting criteria for weighing the potential benefits and costs of grey literature searches.

Outline of Part B

Despite the excellent suggestions made in Chapters 3 and 4, it is safe to say that publication bias will remain a problem in many disciplines for the foreseeable future. The second part of this book presents several statistical methods that have been developed to identify, quantify and assess the impact of publication bias on meta-analyses. Essentially three kinds of techniques have been developed to help analysts deal with publication bias in meta-analysis. One set of techniques is designed to detect publication bias. This set of techniques includes graphical diagnostics such as the funnel plot and explicit statistical tests for the statistical significance of publication bias. In Chapter 5, Jonathan Sterne, Betsy J. Becker and Matthias Egger define funnel plots and demonstrate how they can be used as a graphical tool to for the investigation of publication bias. Specific issues discussed in this chapter are the effects of choice of axes for these plots, and the need to consider explanations in addition to publication bias for funnel plot asymmetry. In Chapter 6, Jonathan Sterne and Matthias Egger describe and illustrate two statistical tests for funnel plot asymmetry, the Begg and Mazumdar (1994) non-parametric test based on the rank correlation between intervention effect estimates and their sampling variances, and the Egger *et al.* (1997) regression method, which tests for a linear association between the intervention effect and its standard error. Sterne and Egger also provide information about the power of these tests, and caution against their use when a meta-analysis contains only a small number of studies.

The second set of techniques is designed to assess the sensitivity of conclusions of an analysis to the possible presence of publication bias. Chapter 7, by Betsy J. Becker, describes and illustrates the first statistical method developed for the assessment of publication bias, file-drawer analysis, originally described by Robert

Rosenthal (1979). This method has been commonly referred to as the failsafe *N*, a term coined by Harris Cooper (1979). Becker also reviews other approaches to file-drawer analysis, including one that was intended to overcome some of the limitations of the original. Her conclusion is that all of these methods should be abandoned in favour of the more sophisticated methods described in the other chapters in this part of the book.

The third set of techniques is designed to adjust estimates for the possible effects of publication bias under some explicit model of publication selection. In Chapter 8, Sue Duval describes trim and fill, a method that she and the late Richard Tweedie developed to estimate and adjust for the number of missing studies (due to publication bias) in a meta-analysis. The trim and fill method is illustrated using a detailed worked example, in addition to its application to the three common data sets used throughout Part B. Chapter 9, by Larry Hedges and Jack Vevea, explicates the general selection model approach to the assessment of publication bias, and demonstrates how it is used to correct for bias. The authors show how their method, as well as that of John Copas, can be used to detect and correct for bias in the three common data sets used throughout the book. As the reader will see, trim and fill is relatively simple to implement and involves relatively little computation, while the Hedges–Vevea and Copas methods involve considerable computation.

In Chapter 10, Alex Sutton summarizes the results of empirical investigations that have been carried out to assess the extent of publication bias present in various scientific literatures, using the methods described earlier in Part B. He concludes on the basis of these results that publication bias assessment should become a routine part of every meta-analysis. In Chapter 11, Michael Borenstein discusses a number of computer programs that can be used to address publication bias, and shows how these would be used to apply the statistical procedures discussed throughout this volume. This chapter should be of great practical value for any researcher who wishes to investigate whether publication bias is likely to be a problem in his or her meta-analysis.

While Part B of the book is necessarily more technical then the other parts, we believe that it is generally accessible to the non-statistically minded reader. Particularly technical sections of these chapters, which can be skipped by those who are not interested in statistical fine points without loss of continuity, are identified with an asterisk.

Outline of Part C

The final part of the book describes advanced and emerging issues related to publication bias. Chapter 12, by Alex Sutton and Terri Pigott, provides a taxonomy of types of missing data. Sutton and Pigott describe and evaluate the application of standard missing-data imputation techniques to meta-analysis, and outline the need for the development of new methods in this area. Chapter 13, by Tom Trikalinos and John Ioannidis, considers how treatment effects in medicine evolve over time and the impact that selective publication may have on such evolution. In Chapter 14, Lesley Stewart, Jayne Tierney and Sarah Burdett discuss the advantages of obtaining individual participant data, rather than relying on published summary results, in

combating publication bias. They suggest that high-quality individual participant data meta-analyses may offer a 'gold standard' for research synthesis, at least in the domain of randomized controlled trials. One of the difficulties in trying to diagnose whether publication bias is present in a meta-analysis data set is that the influence of other factors may mimic the appearance of publication bias. In Chapter 15, John Ioannidis considers how to attack the difficult task of disentangling such factors from true publication bias. In Chapter 16, Scott Halpern and Jesse Berlin reflect on data suppression that may occur for other reasons than those traditionally considered to cause publication bias. These factors include the financial, political, ideological and professional competing interests of investigators, research sponsors, journal editors and other parties. Notable events in the pharmaceutical industry, which we mentioned at the beginning of this chapter, and that received much attention as this book was being completed, suggest that these issues merit serious attention from research synthesists.

OUR MODEST PROPOSAL

We hope that reading this book will convince our audience that is imperative for every meta-analysis to include an analysis of publication bias and that this should be reported as a standard part of the results. We suggest that such reports focus on the practical impact of publication bias. To discuss the practical impact of publication bias it helps to consider three levels of impact, based on the concordance between (1) the results that *are* reported and (2) our best guess (informed by the results of our publication bias analyses) of what the results might look like if all relevant studies had been included. The impact of bias could be called 'minimal' when the two versions of the analysis would yield essentially similar estimates of the effect size. The impact could be considered 'modest' when the effect size would change substantially but the key finding (that the treatment is or is not effective) would remain in force. The impact could be labelled 'severe' when the basic conclusion of the analysis (for example, that the treatment is clinically useful, or that it is not) is called into question. The surveys on this topic, as reviewed by Sutton, in Chapter 10, suggest that publication bias exists in most published meta-analyses, but that the conclusions are nevertheless valid in most cases. In the meta-analyses surveyed so far, the impact of bias is minimal in approximately 50 %, modest in about 45 %, and severe in only 5 % of the analyses surveyed. It also appears to be the case that the amount of bias varies substantially between fields of research. For example, we suspect that publication bias may be more likely in the social sciences, which are characterized by many small and isolated studies, than in medicine, where studies are more likely to be larger, better funded, and better publicized. The prevalence of bias will also likely vary with the experience and resources of the researchers conducting the meta-analysis. The bias cited in the surveys reviewed by Sutton is based primarily on meta-analyses from the Cochrane database, whose researchers are trained to do extensive searches, and which typically include some 30 % more studies than meta-analyses on the same topic that appear in journals. Therefore, the bias cited in these surveys is probably less than the bias one would expect in other fields.

In cases where publication bias analyses suggest that severe bias may exist, this can serve as a warning to researchers and practitioners to regard the initial results cautiously, and to avoid potentially serious mistakes such as recommending an intervention or policy that could be useless or even harmful. Based on the existing state of knowledge in the field, we are hopeful that, in the majority of cases, the publication bias analysis will show that bias probably had little impact. This is also critically important information, as it allows us to have confidence that the meta-analysis is valid.

Finally, we note that it is important to address bias not only to ensure the integrity of the individual meta-analysis, but also to ensure the integrity of the field. When a meta-analysis is published that ignores the potential for bias and is later found to be incorrect, the perception is fostered among editors and researchers that meta-analyses cannot be trusted. By encouraging the prevention and assessment of, and adjustments for, publication bias, we hope to further the use and usefulness of meta-analysis.

REFERENCES

Begg, C.B. and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, **50**, 1088–1101.

Cooper, H.M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology*, **37**, 131–146.

Cooper, H.M. (1998). *Synthesizing Research: A Guide for Literature Reviews*, 3rd edition. Thousand Oaks, CA: Sage.

Egger, M., Davey Smith, G. and Altman, D. G. (2000). *Systematic Reviews in Health Care: Meta-analysis in Context*. London: BMJ Books.

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, **315**, 629–634.

Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin*, **86**, 638–641.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Song, F., Easterwood, A., Gilbody, S., Duley, L. and Sutton, A.J. (2000) Publication and other selection biases in systematic reviews. *Health Technology Assessment*, **4**(10).