Journal Quality, Effect Size and Publication Bias in Meta-analysis

Paul A. Murtaugh

NRCSE

Technical Report Series

NRCSE-TRS No. 068

July 25, 2001

The NRCSE was established in 1997 through a cooperative agreement with the United States Environmental Protection Agency which provides the Center's primary funding.

Journal Quality, Effect Size and Publication Bias in Meta-analysis

Paul A. Murtaugh

Department of Statistics, Oregon State University

Corvallis, Oregon 97331, U.S.A.

e-mail: murtaugh@stat.orst.edu

Abstract. A simple, empirical method of detecting evidence of publication bias in meta-analyses is proposed, based on the relationship between the strength of the results in published studies and the quality of the journals in which they appear. In an illustration of its use, the method is applied to published meta-analyses of terrestrial plant competition, predation in streams, woody plant growth under elevated CO_2, and marine nutrient enrichment experiments. Statistically significant associations of effect strength and journal quality were found in two of the four meta-analyses.

Keywords: journal impact factor; log response ratio; meta-analysis; publication bias; research synthesis

Introduction

Meta-analysis, the quantitative synthesis of results from multiple studies of a single scientific phenomenon, is being increasingly used in ecology and many other fields (Rosenthal 1991; Egger and Davey Smith 1997; Osenberg, Sarnelle and Goldberg 1999). The preferential publication of studies with clear-cut or compelling results — because of selective submission of manuscripts by investigators and/or selective acceptance by journals — may introduce an important bias in meta-analyses based strictly on published work. Evidence of this "publication bias" has been gathered with a variety of methods (Light and Pillemer 1984; Begg and Berlin 1988; Dear and Begg 1992; Begg 1994; Begg and Mazumdar 1994; Sterling, Rosenbaum and Weinkam 1995; Palmer 1999), and several approaches have been suggested for altering the execution or interpretation of meta-analyses in the face of publication bias (Rosenthal 1979; Begg and Berlin 1988; Begg 1994; Vevea and Hedges 1995; Gleser and Olkin 1996; Duval and Tweedie 2000).

I introduce a simple, empirical method of looking for publication bias in a set of studies contemplated for meta-analysis, which is based on the strength of the studies' results and the quality of the journals in which they appear. I then use the method to look for evidence of publication bias in a set of ecological meta-analyses.

Effect size vs. journal quality

If the magnitude and statistical significance of an estimated effect influence the likelihood that a study's results will be published, then one might expect the strength of the results from published studies to increase with the quality or selectivity of the journals in which they appear. Papers describing studies with weak or inconclusive results may not be submitted to high-profile journals, or, if they are, they may tend to lose out to studies with more clear-cut results in the competition for space in those journals.

Effect strength can be expressed as the estimated magnitude of the effect, the estimate divided by its standard error, or a P-value reflecting the strength of evidence against some null-hypothesized value of the effect. It seems reasonable to suppose that large departures from null-hypothesized values in either direction might attract interest and, therefore, increase the chances of submission or publication of a manuscript. Consequently, I focus on the absolute values of responses in all of the analyses that follow.

A possible measure of journal quality is the frequency with which its articles are cited. In its Journal Citation Reports, the Institute for Scientific Information (1999) includes an "impact factor" for each journal. For each of the articles published in a particular journal over the preceding two years, the number of citations of the article during the Journal Citation

Reports cover year is determined. The average of these citation counts (over articles) is the impact factor for that journal.

A scatterplot of an estimate of effect strength vs. impact factor gives a visual idea of whether higher-quality journals tend to publish papers with stronger results. If such a relationship exists, it seems reasonable to infer that unpublished studies will have the weakest results of all, and that surveys of published papers will therefore yield biased estimates of effect strength.

Examples

I illustrate this approach using a set of ecological meta-analyses described in Table **??**, from a special section of *Ecology* (Downing, Osenberg and Sarnelle 1999; Englund, Sarnelle and Cooper 1999; Goldberg et al. 1999; Hedges, Gurevitch and Curtis 1999). In each case, the results of an experiment were summarized as the log of the ratio of responses measured under two treatments, possibly scaled by time. In my analyses, I used the absolute values of the log response ratios, reasoning that extreme values in either direction would tend to draw attention to a study. See the Appendix for further details about the data.

Individual studies (i.e., published papers) in these meta-analyses often encompassed more than one experiment. The authors treated experiments as their units of analysis, assuming independence of results within and across studies. Because there is evidence of study-to-study variation in the responses, and because my method supposes that it is the extremeness of the whole set of results in a study that influences its likelihood of being published, I chose to use studies as the units of my analyses. I summarized each study as the mean of the results from experiments in that study, and then did weighted least-squares regression of mean responses vs. log-transformed impact factors, with weights equal to the numbers of experiments per study. Other more robust summaries are possible (e.g., nonparametric correlation coefficients), but these tend to be less powerful and less able to incorporate study-specific weights than the regression approach.

Since the statistical significance, as well as the magnitude, of an estimated effect presumably influences the impression it makes on authors and editors, I analyzed standardized estimates (i.e., estimates divided by their estimated standard errors) for the two references that provided the needed information. Variance estimates were directly available for a subset of the studies from Reference 2, and they could be calculated from information in the archived data for Reference 3 (see Hedges et al. 1999 for details). The standardized effect estimates should be inversely related to P-values from hypothesis tests (and to the widths of confidence intervals).

Table 2 shows the results of the regression analyses, and Figure 1 is a graphical representation of two extreme cases. There is statistically significant evidence of an association between strength of results and journal quality for two of the four references. Whether significant or not, the slopes of the regressions of effect strength vs. journal quality are positive in 9 of the 10 analyses reported in Table 2.

If statistical significance is more important than the magnitude of an effect in influencing a study's likelihood of being published, one would expect the relationship of standardized estimates to journal impact factors to be stronger than that of unstandardized estimates. For both subsets of data from Reference 2, the regression slope is more positive when standardized estimates are used. (The P-values are not directly comparable because of the different sample sizes). This pattern does not hold for the plant growth experiments from Reference 3.

It is difficult to know how to summarize the aggregated results in Table 2, given the uncertainty about how best to measure the extremeness of results from individual studies, the unknown extent to which journal impact factor accurately reflects journal quality, the multitude of possible ways of analyzing the relationship between the summaries of effect strength and journal quality, and the interdependence of multiple hypothesis tests performed on data from a single reference. But it is certainly suggestive that most of the relationships of effect strength and journal quality are positive, with statistical significance established in two cases.

It seems likely that studies with powerful designs, careful execution and substantial replication will yield more precise results, and smaller P-values, than less exemplary studies. It is conceivable that the quality of these studies, rather than the strength of their results, is what causes them to appear in high-profile journals — so that an inverse relationship between P-values and journal quality is not necessarily evidence of a bias that should be corrected for. The strongest associations in Table 2, however, come from analyses of the magnitude, rather than the statistical significance, of estimated effects, and it is difficult to see why sloppy studies should tend to yield smaller estimates. Nevertheless, information on the quality of studies could be helpful in the interpretation of the kinds of associations reported in Table 2.

Conclusions

These analyses suggest that, in some meta-analyses, studies showing weak or inconclusive results are less likely to appear in high-profile journals than are studies with more clear-cut results. It seems reasonable to infer in these cases that results from unpublished studies will generally be weaker still, and that meta-analyses based on published work will therefore tend to give biased estimates of effect sizes. To counteract this bias, one could make a renewed effort to find unpublished or obscurely documented studies, or attempt to modify the analysis or the interpretation of its results to account for publication bias (Rosenthal 1979; Begg and Berlin 1988; Begg 1994; Vevea and Hedges 1995; Gleser and Olkin 1996; Duval and Tweedie 2000).

If, on the other hand, a large meta-analysis spanning a variety of journals shows no evidence of an association between response strength and journal impact (like Reference 3 in Table 2), then one might have increased confidence that the analysis has not been seriously tainted by publication bias.

Acknowledgments

I thank A.R. Palmer and an anonymous reviewer for constructive criticisms of an earlier version of the manuscript.

Literature Cited

Begg, C. B. 1994. Publication bias. Pages 399-409 *in* H. Cooper and L. V. Hedges, editors. The handbook of research synthesis. Russell Sage Foundation, New York.

Begg, C. B., and J. A. Berlin. 1988. Publication bias: a problem in interpreting medical data. Journal of the Royal Statistical Society, Series A **151**: 419-463.

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

Dear, K. B. G., and C. B. Begg. 1992. An approach for assessing publication bias prior to performing a meta-analysis. Statistical Science **7**: 237-245.

Downing, J. A., C. W. Osenberg, and O. Sarnelle. 1999. Meta-analysis of marine nutrient-enrichment experiments: Variation in the magnitude of nutrient limitation. Ecology **80**: 1157-1167.

Duval, S., and R. Tweedie. 2000. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. Journal of the American Statistical Association **95**: 89-98.

Egger, M., and G. Davey Smith. 1997. Meta-analysis: Potentials and promise. British Medical Journal **315**: 1371-1374.

Englund, G., O. Sarnelle, and S. D. Cooper. 1999. The importance of data-selection criteria: Meta-analyses of stream predation experiments. Ecology **80**: 1132-1141.

Fisher, R. A. 1970. Statistical methods for research workers. Hafner, New York.

Gleser, L. J., and I. Olkin. 1996. Models for estimating the number of unpublished studies. Statistics in Medicine **15**: 2493-2507.

Goldberg, D. E., T. Rajaniemi, J. Gurevitch, and A. Stewart-Oaten. 1999. Empirical approaches to quantifying interaction intensity: Competition and facilitation along productivity gradients. Ecology **80**: 1118-1131.

Hedges, L. V., J. Gurevitch, and P. S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology. *Ecology* **80**: 1150-1156.

Institute for Scientific Information. 1999. 1998 Journal Citation Reports On Microfiche, Science Edition. Philadelphia.

Light, R. J., and D. B. Pillemer. 1984. Summing Up: The science of reviewing research. Harvard University Press, Cambridge.

Osenberg, C. W., O. Sarnelle, S. D. Cooper, and R. D. Holt. 1999. Resolving ecological questions through meta-analysis: Goals, metrics, and models. Ecology **80**: 1105-1117.

Osenberg, C. W., O. Sarnelle, and D. E. Goldberg. 1999. Meta-analysis in ecology: Concepts, statistics, and applications. Ecology **80**: 1103-1104.

Palmer, A. R. 1999. Detecting publication bias in meta-analyses: A case study of fluctuating asymmetry and sexual selection. American Naturalist **154**: 220-233.

Rosenthal, R. 1979. The “file drawer problem” and tolerance for null results. Psychological Bulletin **86**: 638-641.

Rosenthal, R. 1991. Meta-analytic procedures for social research. Sage, Beverly Hills.

Sterling, T. D., W. L. Rosenbaum, and J. J. Weinkam. 1995. Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. The American Statistician **49**: 108-112.

Vevea, J. L. and L. V. Hedges. 1995. A general linear model for estimating effect size in the presence of publication bias. Psychometrika **60**: 419-435.

Appendix

The data from the special section of *Ecology* are archived at http://esa.sdsc.edu/Archive/. Two other meta-analyses from this issue, in the paper by Osenberg, Sarnelle, Cooper and Holt (1999), were not included because only two journals were represented (snail data), or there was no clear null hypothesis (mutual interference data).

One study (Reice 1983) was omitted from Reference 2, and one study (Couteaux et al. 1992) was omitted from Reference 3, because they were from books, not journals. In Reference 4, one study (Fisher et al. 1992) was omitted because no journal reference was given, and two studies (Le Rouzic and Bertru 1989, and Thomas et al. 1974) were omitted because their journals were not included in the Journal Citation Reports (ISI 1999).

Reference number	Subject matter	Quantity estimated
1	Terrestrial plant competition experiments	Log of ratio of plant biomasses or survival probabilities
2	Stream predation experiments	Log of ratio of total prey densities
3	Elevated CO_2 and woody plant growth	Log of ratio of plant biomasses
4	Marine nutrient-enrichment experiments	Log of ratio of phytoplankton biomasses, scaled by time

Table 1: Description of the four meta-analyses: (1) Goldberg et al. 1999, (2) Englund et al. 1999, (3) Hedges et al. 1999, and (4) Downing et al. 1999. In each case, the "null" value of the response is zero.

Ref.	Subset of data	No. of expts.	No. of studies	Response vs. log of journal impact: Int.	Slope	*P*-value
1	Biomass	86	7	−0.119	1.455	0.173
	Survival	28	6	−0.117	0.427	0.313
2	Fish predators, total prey					
	Unstandardized *	45	28	0.264	**0.189**	0.006
	Standardized	33	18	1.384	0.587	0.103
	Invert. predators, total prey					
	Unstandardized	38	12	0.258	0.102	0.409
	Standardized	11	5	1.207	0.639	0.307
3	All					
	Unstandardized	101	28	0.289	0.027	0.514
	Standardized	101	28	3.421	−0.323	0.727
4	Nitrogen experiments	137	11	0.149	**0.065**	0.003
	Phosphorus experiments *	103	10	0.028	0.014	0.334

Table 2: Results of weighted least-squares regressions of study-specific means of the absolute values of the log response ratios vs. log of journal impact factors, for various responses and data subsets. Studies are the units of analysis, weighted by the number of experiments per study. Slopes for which $P < 0.05$ are in boldface. Asterisks denote results that are plotted in Figure 1.

Figure Legend

Figure 1.

Absolute value of log response ratio vs. log of journal impact factor, for studies of (a) the effects of fish predators on invertebrate prey in streams (Englund et al. 1999), and (b) the effects of elevated CO_2 on woody plant growth (Hedges et al. 1999). Experiments from the same study are indicated by the same plotting character, slightly jittered for visibility. The lines are from weighted least-squares regressions.

Figure 1