Political Analysis, 8:2

Publication Bias Reconsidered

Lee Sigelman
The George Washington University

In political science and many other disciplines, statistically significant results—rejections of the null hypothesis—are achieved more frequently in published than in unpublished studies. Such "publication bias" is generally seen as the consequence of a widespread prejudice against statistically nonsignificant results. I argue that evidence of such a prejudice is in surprisingly short supply and that publication bias can occur even in the absence of such a prejudice and even if the review process is functioning perfectly. More importantly, publication bias may stem from dutiful application of standards of scientific inquiry rather than from irrational prejudice.

1 Introduction

IN A PERFECTLY random world, nothing explains anything, and the null hypothesis (H_0) rules. In a perfectly ordered world, everything can be explained, and H_0 is an outcast. To judge from the research findings that are regularly reported in the leading journals of political science, the world of politics must be a very orderly place, indeed, and political scientists must be masters of exposing its regularities. In the face of what critics castigate as the discipline's weak theoretical foundations, mushy concepts, and unreliable measures, political scientists manage with gratifying regularity to uncover statistically significant patterns in their data—deviations from what would be expected based solely on chance. Consider, for example, a recent (1997) volume of the *American Journal of Political Science*. That volume contains, by my count, 3064 tests of H_0. Purely by chance, H_0 would be rejected at the .05 confidence level in 5% of these tests, i.e., 153 times. In fact, though, it was rejected more than 10 times that often: 54.7% of the reported tests, or 1676 of the 3064, were statistically significant.[1]

The poor performance of H_0 is by no means unique to political science. For example, sociologists appear to be even more successful than political scientists at achieving statistical significance, to judge from the research published in the 1997 volume of the *American Sociological Review*. In that volume, 62.3% of the reported tests of H_0 yielded statistically significant results. In many other research fields the frequency at which H_0 is rejected is even higher, routinely topping 90% and sometimes even approaching 100% (Sterling 1959; Sterling et al. 1995). Rather than being masters of exposing empirical regularities, then, political scientists may actually be laggards.

Author's note: Thanks are due to Robert Goldfarb, Gary King, Jim Lebovic, Forrest Maltzman, Carol Sigelman, Jack Wright, and Langche Zeng for their helpful comments and suggestions.

[1] This count is based on coefficients presented in tables and excludes significance tests involving the constant term in regression equations.

Copyright 1999 by the Society for Political Methodology

But are the world in general and politics in particular really as orderly as a reading of published research studies seems to suggest? Probably not, for it is widely understood that statistically significant results—rejections of H_0—are achieved more frequently in published than in unpublished studies, a phenomenon known as "publication bias" (Begg and Berlin 1988; Card and Krueger 1995; Coursol and Wagner 1986; Dickersin 1990; Dickersin and Min 1993; Dickersin et al. 1992) or "the file-drawer problem" (Rosenthal 1979, 1991)—the latter to acknowledge the purported tendency of statistically nonsignificant results to languish in the researcher's files because they are deemed unpublishable.

If statistically nonsignificant results tend to end up in file drawers, then the obvious danger is that what we read in the journals may bear only the slightest resemblance to what occurs in the real world. Imagine that 20 researchers, working independently, investigate a phenomenon that is truly random. Purely by chance, 1 of the 20 should observe a statistically significant relationship ($p < .05$). The 19 unlucky researchers who have nothing significant to report presumably either will be so discouraged by their "failure" that they shift their attention to something else or will write up their "negative" results nonetheless, only to have them rejected by reviewers and editors who crave significant findings. In contrast, the lucky one who has something "positive" to report will encounter much less difficulty in getting these findings published. The problem, of course, is that the 19 sets of findings that never make it into the public domain are the correct ones, while the findings that make it into print do so only because the author has committed the Type I error of rejecting H_0 when it is true.

In fields in which meta-analytic literature reviews are common, such as psychology, some innovative techniques have been devised for coping with publication bias. Although this phenomenon is less widely recognized in fields that rely more heavily on traditional text-based literature reviews, such as political science [see Lau et al. (1999) for a rare exception], it is no less problematic in those fields. If anything, it is more problematic, precisely because it is less widely recognized.

A good illustration of these problems is provided by the Bayesian "model-averaging" technique that Bartels (1997) recently introduced as a tool for analyzing political data. Model-averaging is intended as a corrective to the tendency of data analysts to report results from several alternative specifications of a model, a situation for which classical statistical theory is inappropriate. To take specification uncertainty into account, results are averaged across the alternative model specifications that a researcher reports. This procedure assumes, of course, that the alternative specifications that need to be taken into account in model-averaging are the ones that are actually reported. However, experience suggests that data analysts are likely to try out numerous specifications before settling on the relative few that they actually report and, further, that the specifications most likely to go unreported are those that do not "work." If this is so, then publication bias undermines model-averaging, for results that are unreported cannot be averaged in with the rest.

What accounts for publication bias? The conventional and seemingly inescapable explanation is that the research process is permeated, from start to finish, by a prejudice against statistically nonsignificant results, a pervasive sense that not producing statistically significant results constitutes "failure." If researchers, reviewers, and editors alike, or any subset thereof, believe that precious journal space should be conserved for reports of "positive" findings, then publication bias seems both inevitable and problematic—inevitable if reviewers are reluctant to recommend, editors to accept, and authors even to submit research studies that fail to reject H_0 and problematic because, in that case, the findings reported in published studies do not accurately represent the findings obtained in all the studies that have been conducted. That is, publication bias is seen as a pressing problem because it deters us from understanding the world as it really is—the core goal of inquiry.

That, at least, is the conventional view, and it is the view with which I take issue in what follows. The idea that prejudice against statistically nonsignificant findings pervades the scientific community provides a simple, straightforward, and powerful explanation of publication bias. Even so, I argue that such a prejudice is not a necessary condition of publication bias. That is, publication bias—the overrepresentation of statistically significant findings in published studies—can occur even in the absence of any prejudice against failure to reject H_0 and even if the review process is functioning perfectly. Am I therefore going to deny something that seems undeniable: the existence of prejudice against statistically nonsignificant findings? No, though I know of no solid evidence that such prejudice actually exists. Nor do I deny that publication bias stems, in part, from such prejudice. However, I question whether publication bias is as widespread as it is generally believed to be. Much more importantly, I contend that publication bias stems, in part, from an altogether different sort of prejudice—one that is anything but pernicious or unscientific and which bolsters rather than undermines confidence in published research findings. That is, I argue that, in addition to being a product of prejudice, publication bias is also a consequence (albeit an unintended one) of the evolutionary process whereby methodologically deficient research is weeded out and of the survival of the fittest research resulting from the application of rigorous scientific standards.

It is not my purpose here to enter the long-standing but recently reignited controversy over the logic and value of null hypothesis significance testing, especially given Gill's (1999) excellent overview of this controversy, to which I have nothing to add. Whether political scientists should be engaging in so much significance testing against H_0 in the first place (more than 3000 such tests in a single volume of a single journal!) is a good question, but it is not my question. Given that political scientists are such heavy users of null hypothesis significance testing, my question is whether publication bias is as problematic as it is generally assumed to be.

2 Bias as an Unrepresentative Outcome and as a Prejudiced Attitude

I begin by distinguishing between two meanings of "bias." Bias in one sense is an outcome or a result. When we refer to a statistical estimate as biased, we mean that its expected value is not equal to the true underlying population parameter. If statistically significant results are more likely to be achieved in published studies than in all studies, unpublished as well as published, then the findings presented in published studies comprise an unrepresentative sample of all extant findings. In another sense, though, bias is an attitude or predisposition, a prejudice rather than a result. If reviewers, editors, or authors consider research unworthy because it fails to reject H_0, then we can speak of such prejudice as a bias that is conceptually distinct from, though it certainly could lead to, bias in the outcome sense.

Unfortunately, the distinction between bias-as-outcome and bias-as-attitude is routinely overlooked in studies of publication bias.[2] What is actually measured in such studies is bias-as-outcome, the higher rate of rejecting H_0 in published than in unpublished studies. What is then inferred is bias-as-attitude, the operation of prejudice against failure to reject H_0. That is, biased outcomes are taken as prima facie evidence of biased attitudes, which are then invoked, circularly, as the explanation of biased outcomes.

By distinguishing between bias-as-outcome and bias-as-attitude, I treat the link between the two as empirical rather than definitional. This means that although bias-as-attitude provides a parsimonious explanation of bias-as-outcome, there is no certain link between them.

[2] Though not invariably; see Begg and Berlin (1988, pp. 422–423).

Even if prejudice against statistically nonsignificant results were widespread, it could be overridden elsewhere in the process, in which case it would be insufficient to produce the oft-observed outpouring of statistically significant results. Similarly, even when an outpouring of statistically significant results occurs, it might (as I argue below) be traceable to sources other than prejudice against statistically nonsignificant results, and—most importantly—it might not be a cause for concern.

In sum, distinguishing between bias-as-outcome and bias-as-attitude has the effect of decoupling two phenomena that are often, and in my view mistakenly, treated as one. The implications of this decoupling are twofold: prejudice against statistically nonsignificant results must be established independently rather than inferred from the differential rate at which H_0 is rejected in published and unpublished studies; and the empirical link between the two phenomena must be demonstrated rather than assumed.

3 Evidence of Bias-As-Attitude

How convincing is existing evidence of the operation of widespread prejudice against statistically nonsignificant results? Those who have studied publication bias have followed three routes to establishing the operation of prejudice against statistically nonsignificant results: assumption, usually implicit and usually based on the circular reasoning described above; appeal to "common knowledge" or recounting of an anecdote or two; and reference to systematic evidence. Of these, the first two are obviously problematic, but so, surprisingly, is the third.

Insofar as I can determine, the only systematic evidence that is ever cited concerning prejudice against statistically nonsignificant results comes from a study conducted by Mahoney (1977). There is an obvious irony in the fact that a literature in which the virtues of replication are so often extolled rests so decisively on the results of a single study. What is even more striking is that the Mahoney study does not even address the issue of prejudice against statistically nonsignificant results, let alone document such prejudice.

In the study in question, Mahoney (1977) delved into what he called "confirmatory bias," the tendency to credit evidence that supports one's a priori beliefs and to discredit disconfirmatory evidence. To assess that tendency, he designed an experiment in which reviewers who were exponents of a certain theoretical perspective were asked to assess a research report in which the results were either consistent with their perspective, inconsistent with their perspective, or mixed. The experimental results indicated that reviewers were indeed more favorably disposed to reports of findings that were consistent with their own perspective than to reports of findings that were inconsistent with their own perspective or were only partially supportive thereof. These results were clear-cut. Even so, they did not establish the operation of prejudice against statistically nonsignificant findings, nor did Mahoney claim that they did.

In the first place, recall that Mahoney was testing for "confirmatory bias." Prejudice in favor of confirming one's preconceptions has some elements in common with prejudice in favor of statistically significant results, but they are not the same thing, as a simple example demonstrates. Suppose that an adherent of Theory X were reviewing a paper that reported a statistically significant effect inconsistent with Theory X but consistent with rival Theory Y. A prejudice in favor of statistically significant results would predispose the reviewer toward a favorable assessment of the paper, but a prejudice in favor of confirming one's own preconceptions would predispose the same reviewer toward an unfavorable assessment. Clearly, establishing the operation of prejudice in favor of confirming one's preconceptions is not equivalent to establishing the operation of prejudice in favor of rejecting H_0.

Even more problematically, Mahoney's experiment did not even involve statistical significance or nonsignificance. In the research reports he sent to reviewers, no statistical tests were presented and no mention was made one way or the other of the statistical significance of the results described therein. Because statistical significance or nonsignificance played no part in Mahoney's experiment, his findings cannot credibly be interpreted as evidence of bias against statistically nonsignificant results.

The best evidence of prejudice against statistically nonsignificant results, then, is in fact no evidence at all. Does this mean that there is no such prejudice? Obviously not, but it does mean that evidence of such a prejudice is in remarkably short supply.

4 Evidence of Bias-As-Outcome

By comparison, evidence of publication bias-as-outcome is relatively abundant. For example, Coursol and Wagner (1986) found that researchers submitted for publication 82% of papers that reported a positive outcome but only 43% of papers that reported a negative outcome; and 80% of the positive submissions were accepted, but only 50% of the negative submissions. If we take such estimates at face value, then we can get a good sense of the magnitude of publication bias-as-outcome.

But should we take such estimates at face value? Precisely to the extent that we take publication bias-as-outcome seriously, we must also take seriously the possibility that published estimates of its frequency are themselves beset by publication bias. If they are, then biased outcomes cannot be nearly as serious a problem as published estimates such as those cited in the previous paragraph seem to indicate. Paradoxically, taking seriously the idea of biased outcomes undermines our confidence that outcomes are as biased as they are widely believed to be. No less paradoxically, only if we discount the very idea of publication bias and assume instead that published estimates are accurate can we confidently accept published estimates of the frequency of biased outcomes. In this sense, the idea of publication bias poses a self-defeating prophecy: if we accept it, we must apply it to itself, in which case we become less inclined to accept it.

5 Sources of Publication Bias-As-Outcome

I have now decoupled bias-as-attitude from bias-as-outcome and shown that the evidence of the latter, like that of the former, is problematic. Rather than leaving matters at this point and calling for new and more convincing data on both forms of bias, let me now move my argument forward by laying these data-centered problems aside, suspending disbelief, and taking published estimates of publication bias-as-outcome at face value. The issue then becomes how to account for bias-as-outcome, which I am now stipulating for the sake of argument. The possibility that immediately springs to mind, of course, is that bias-as-outcome is a consequence of bias-as-prejudice. Without denying that possibility, I now point the finger of suspicion at other potential sources of publication bias-as-outcome. More specifically, I argue that even a totally prejudice-free selection process, by which I mean a process in which judgments are based solely on scientific merit, should often result in outcomes biased in favor of statistically significant results.

The primary purpose of the process that begins when researchers first plan their studies and ends when papers are accepted or rejected for publication is to produce trustworthy research findings and place them in the public domain. Assume that this is what actually happens, i.e., that the process works as it should, so that the studies that are undertaken, completed, submitted, and ultimately published are more meritorious, according to widely accepted canons of scientific research, than those that never appear in print. We need not

assume that every published study is better than every unpublished one, just that, on average, the former are better than the latter. The result would be an outcome biased toward higher quality in published research than in unpublished research—a publication bias of a different sort than the one toward statistical significance and an outcome to which no one (with the probable exception of authors of studies that do not get published) would object.

Regrettably, much research is less than fully meritorious from the perspective of research design. There is no end to the ways in which research can fall short; one is reminded of Tolstoy's opening sentence in *Anna Karenina*: "All happy families are alike but an unhappy family is unhappy after its own fashion." Still, it is instructive to review a few of these imperfections and to consider their implications.

Inadequate sample size. Large samples are, ceteris paribus, preferable to small ones. At a given level of confidence, the larger the sample, the more accurate the estimate of the true parameter; for a given estimate of the true parameter, the larger the sample, the greater the confidence one can place in the estimate. Accordingly, an estimate that falls short of statistical significance at a conventional level (say, $p < .05$) in a small-sample study may well prove statistically significant in a large-sample study; even a tiny estimated effect can be statistically significant if the sample is sufficiently large. The implication is that weeding out small-sample analyses, which is often defensible on methodological grounds, should produce a bias toward statistical significance among the surviving studies, for the larger the sample, the easier it becomes to reject H_0. Here, then, we see that an outcome biased toward statistical significance can result, not just from prejudice against statistically nonsignificant findings, but also from rejection of research that is poorly designed—in this case, because of inadequate sample size. Of course, large-sample studies may lead researchers to reject H_0 on the basis of statistically significant but substantively trivial results; that, however, is an argument against overreliance on significance testing, not against use of large samples.

Unreliable measurement. Another common design problem is measurement unreliability, which appears in different guises in different research contexts. As a simple example, suppose that X is truly correlated with Y. To measure X and Y, we use two sets of response items, from which we create composite X and Y scales. If the correlations between items that compose a given scale are modest, then the X and Y scales will each have a large "noise" (random error) component. Because unreliable measures obscure true relationships, it is appropriate, on scientific grounds, to look askance at research that employs measures of marginal or low reliability. The primary consequence of weeding out such studies is that published studies will present more trustworthy estimates of the relationship between X and Y than do unpublished studies—hardly an objectionable outcome.

Of course, not all measurement error is random. However, a secondary consequence of weeding out research beset by random measurement error may be the opening of a gap between published and unpublished studies in the reported correlation between X and Y. Random error can only attenuate the observed correlation between the X and the Y scales: the louder the noise, the closer the observed correlation will be to 0. Holding sample size constant, the closer the observed X–Y correlation is to 0, the less likely it is that H_0 will be rejected. Thus, even in the absence of a prejudice in favor of rejecting H_0, weeding out studies that rely on noisy measures will have the same effect as a prejudice in favor of rejecting H_0: it will open a gap between published and unpublished studies in the frequency at which H_0 is rejected.

Weak theoretical foundation. Some research is not guided by an intelligible set of ideas, let alone a coherent theoretical model. In "exploratory" studies, researchers proceed without clearly defined expectations of what the data will reveal, or with only the flimsiest of hypotheses. In practice, research conducted without benefit of compelling ideas often turns

out to be research without compelling findings; when one conducts a study not expecting much, not much is exactly what one should be prepared to find. This is not to say that empirical work based on strong theoretical foundations invariably produces strong findings, for it does not. Nor is it to deny the possibility that research based on weak theoretical foundations can produce strong findings, as often occurs, for example, in time-series analysis. But it is to say that rejection of H_0 should be more likely when such rejection follows directly from a coherent theoretical model than when there was no strong a priori basis for hypothesizing such rejection in the first place. So research based on a strong theoretical foundation not only is preferable on scientific grounds, but also is more likely to produce statistically significant results. Accordingly, if the selection process displays a preference for studies that have solidly grounded hypotheses, then we should anticipate a gap between published and unpublished studies in the probability of rejecting H_0, irrespective of any prejudice against failure to reject H_0. Moreover, even if theory-based studies are weeded out along with weakly grounded ones, publication bias should be expected if the theoretically oriented studies are rejected because they test hypotheses that have been inappropriately derived from a coherent model, have been appropriately derived from a dubious model, or rest on little prior evidence.[3]

Scientific conservatism. Publication bias can also stem not from defects in research design but from the very conservatism of science. Reluctant to reject the null hypothesis unless the evidence against it is extremely strong, scientists conventionally make it difficult to commit a Type I error. They do this by rejecting H_0 only if a statistical test shows that the observed effect would have occurred by chance no more than 5 times in 100 (or, more extremely, 1 time in 100 or 1 time in 1000) if the true effect were zero.[4] Of course, given data of fixed quality and quantity, the chances of making a Type I error are inversely related to the chances of making the Type II error of accepting H_0 when it is false. It follows that because scientists leave themselves so little chance of committing a Type I error, they leave ample room to commit a Type II error. In this sense, results risking Type I error (i.e., results rejecting H_0) warrant greater trust than results risking Type II error (i.e., results failing to reject H_0)—and it seems highly appropriate that results that warrant greater trust would be deemed more publishable than less trustworthy results.[5]

This brief consideration of inadequate sample size, unreliable measurement, weak theoretical foundation, and scientific conservatism by no means exhausts the inventory of reasons why research studies can be found unworthy of publication. However, even this short list is sufficient to demonstrate that publication bias-as-outcome can stem from sources other than prejudice against statistically nonsignificant findings. These sources themselves reflect prejudices, albeit of an altogether defensible sort: prejudice against research that is methodologically deficient and against findings that are untrustworthy. To the extent that publication bias is a consequence of survival of the fittest research, then such bias strikes me not as a problem that needs to be corrected but rather as an indication that the selection process is working properly. What is worrisome is not publication bias-as-outcome per se, but biased outcomes that are attributable to "unreasonable" prejudice. A predisposition to maintain high methodological standards and to publish trustworthy results does not qualify as unreasonable.

[3]In this sense, research that operates within the framework of an established paradigm, i.e., "normal science" in Kuhn's (1962) sense, should be expected to yield more significant results than other research. However, for a good summary of what can go wrong in testing theoretically derived hypotheses, see Meehl (1990).

[4]On the idiosyncratic origins of the .05 convention, see Cowles and Davis (1982).

[5]I am grateful to Langche Zeng for suggesting this point.

The point is not that "bad" research necessarily produces nonsignificant results, for, as Greenwald (1975, p. 3) has properly noted, "Some . . . very common types of incompetence are much more likely to produce false positive or significant results." Rather, the point is that in many common circumstances, what may outwardly appear to be a bias against nonsignificant findings may actually be a bias for methodological rigor. To return to the distinction with which I began, bias-as-attitude is simply not inferable from bias-as-outcome.

6 Is Prejudice Against Statistically Nonsignificant Results Always Unreasonable?

Having said that, I must add that in my view even a flat-out prejudice against nonsignificant results does not necessarily qualify as unreasonable. As noted above, the conservatism of science makes it rational to place greater trust in studies that reject H_0. Beyond that, consider an exploratory study in which a research hypothesis is set forth without any compelling theoretical foundation. If it turns out that H_0 cannot be rejected, then, because there was little reason to expect a significant effect in the first place, the study will make no real contribution; even if it is otherwise impeccable, there is little reason to publish it. On the other hand, if the very same study does report a significant effect, it will at least have the potential to trigger a scientific advance, for it will pose a riddle (how to account for the observed effect) whose solution might yield valuable new insights. In this scenario, it seems reasonable not to publish the study if it produces nonsignificant results but to publish it if it does reject H_0.[6] I grant that this distinction reflects a prejudice against nonsignificant findings, but I would defend that prejudice as reasonable in the circumstances.

In different circumstances I would argue the opposite case. For example, which study seems more likely to yield new insights—one that replicates a long series of earlier positive tests of a well-established research hypothesis or one that fails to replicate the same oft-reported positive result? Of these two studies, the one that fails to reject H_0 seems to make the greater contribution, because it will cause researchers to think more deeply about a relationship that they had come to take for granted.[7] More generally, philosophers of science have often made the argument that, in principle, failures to reject H_0 convey more information than do positive results, given the falsificationist logic of science. Meehl (1978, p. 822) carries this argument a step farther: "Putting it crudely, if you have enough cases and your measures are not totally unreliable, the null hypothesis will always be falsified, regardless of the truth of the substantive theory"—and positive results therefore convey no information. For all these reasons, I do not want to be interpreted as arguing that prejudice against statistically nonsignificant findings is always rational. But neither is it always irrational.

[6] Simon (1988, p. 459) makes exactly the same argument: "Negative 'exploratory' studies are generally less important than positive studies in providing leads for the discovery of improved treatments. The positive findings, until confirmed, will not be adequate for recommending a treatment to practitioners, but they tell investigators what paths to take and what to evaluate in large clinical trials."

[7] It is possible, however, to put an altogether different spin on the same situation. For example, one might emphasize that a study that fails to reject the null hypothesis in the face of a long procession of prior rejections may have failed by chance, in which case the new null findings would not warrant reconsideration of the underlying relationship. De Long and Lang (1992, p. 1259) advance a more sinuous reverse-spin argument: "[S]tudies that fail to reject their null hypotheses are much more likely to be published when prior work has already strongly established a contrary result. This makes it plausible that papers that fail to reject their null hypotheses survive the refereeing process and get published only if the probability that the null hypotheses they test are false is high, for when the null hypothesis is in fact false, earlier work is most likely to have established the contrary presumption that makes the paper's failure to reject interesting." In other words, studies that fail to reject the null hypothesis are most likely to be published when they are most likely to be incorrect. It seems to follow that such studies are uninformative.

7 Conclusion

I do not seriously doubt the existence of a prejudice against failure to reject H_0. However, it is high time for systematic evidence of such prejudice to be produced; we need findings that parallel those reported by Mahoney (1977) but bear directly on prejudice against statistically nonsignificant results. Nor am I naive enough to believe that studies are published purely on the basis of their scientific merit or lack thereof, and on this point the evidence is strong (e.g., Peters and Ceci 1982). Even so, I remain suspicious of published estimates of publication bias, and it is certainly worth noting that despite the alleged prejudice against H_0, failures to uncover publication bias have occasionally made their way into the published literature (e.g., Bero et al. 1994; Rosenthal and Rubin 1988), though they have been conspicuously ignored in reviews of the publication bias literature (e.g., Begg 1994). Accordingly, publication bias-as-outcome may be less pronounced than is widely assumed. Moreover, even if we take published estimates of publication bias-as-outcome at face value, we must grant that it can stem from sources entirely removed from prejudice against failure to reject H_0—sources that are anything but un- or antiscientific. If this is so, then some portion of bias-as-outcome is a consequence, not of prejudice against statistically nonsignificant results, but of dutiful application of consensually held scientific standards. Nor is prejudice against statistically nonsignificant results necessarily unreasonable. It follows that even if one could isolate the portion of publication bias-as-outcome that is due to prejudice against statistically nonsignificant results, one would have to understand the situational context of such prejudice before one could decide whether the observed bias represents a defect or a triumph of the review process.

Past considerations of how to deal with publication bias have suffered from an insufficiently nuanced understanding of the phenomenon, an inflated sense of its magnitude, an overly restrictive understanding of its sources, and an unquestioned assumption that it is a problem that needs to be corrected. A more appropriate approach would begin with a clear demarcation of what is and what is not meant by publication bias, would treat as problematic only those instances that can be linked to unreasonable prejudice against failure to reject H_0, and would focus remediation efforts on instances in which such a linkage has been established rather than treating publication bias-as-outcome per se as a problem to be rooted out.

References

Bartels, L. M. 1997. "Specification Uncertainty and Model Averaging." *American Journal of Political Science* 41:641–674.

Begg, C. B. 1994. "Publication Bias." In *The Handbook of Research Synthesis*, eds. H. Cooper and L. V. Hedges. New York: Russell Sage Foundation, pp. 400–408.

Begg, C. B., and J. A. Berlin. 1988. "Publication Bias: A Problem in Interpreting Medical Data." *Journal of the Royal Statistical Society, Series A* 151:419–445.

Bero, L. A., S. A. Glantz, and D. Rennie. 1994. "Publication Bias and Public Health Policy on Environmental Tobacco Smoke." *Journal of the American Medical Association* 272:133–136.

Card, D., and A. B. Krueger. 1995. "Time-Series Minimum-Wage Studies: A Meta-Analysis." *American Economic Review* 85:238–243.

Coursol, A., and E. E. Wagner. 1986. "Effect of Positive Findings on Submission and Acceptance Rates: A Note on Meta-Analysis Bias." *Professional Psychology* 17:136–137.

Cowles, M., and C. Davis. 1982. "On the Origins of the .05 Level of Statistical Significance." *American Psychologist* 37:553–558.

De Long, J. B., and K. Lang. 1992. "Are All Economic Hypotheses False?" *Journal of Political Economy* 100:1257–1272.

Dickersin, K. 1990. "The Existence of Publication Bias and Risk Factors for Its Occurrence." *Journal of the American Medical Association* 263:1385–1389.

Dickersin, K., and Y.-I. Min. 1993. "Publication Bias: The Problem that Won't Go Away." *Annals of the New York Academy of Science* 703:135–146.

Dickersin, K., Y.-I. Min, and C. L. Meinert. 1992. "Factors Influencing Publication of Research Results: Follow-up of Applications Submitted to Two Institutional Review Boards." *Journal of the American Medical Association* 267:374–378.

Gill, J. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52:647–674.

Greenwald, A. G. 1975. "Consequences of Prejudice Against the Null Hypothesis." *Psychological Bulletin* 82: 1–20.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lau, R. R., L. Sigelman, C. Heldman, and P. Babbitt. 1999. "The Effects of Negative Political Advertisements: A Meta-Analytic Assessment." *American Political Science Review* (in press).

Mahoney, M. J. 1977. "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System." *Cognitive Therapy and Research* 1:161–175.

Meehl, P. E. 1990. "Why Summaries of Research on Psychological Theories Are Often Uninterpretable." *Psychological Reports* 66:195–244.

Peters, D. P., and S. J. Ceci. 1982. "Peer-Review Practices of Psychological Journals: The Fate of Published Articles Submitted Again." *Behavioral and Brain Sciences* 5:187–255.

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*. Newbury Park, CA: Sage.

Rosenthal, R., and D. B. Rubin. 1988. "Comment: Assumptions and Procedures in the File Drawer Problem." *Statistical Science* 3:120–125.

Simon, R. 1988. "Comment on Begg and Berlin." *Journal of the Royal Statistical Society, Series A* 151:419–459.

Sterling, T. D. 1959. "Publication Decision and the Possible Effects on Inferences Drawn from Tests of Significance—or Vice Versa." *Journal of the American Statistical Association* 54:30–34.

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." *American Statistician* 49:108–112.