Publication Bias in Psychological Science: Prevalence, Methods for Identifying and Controlling, and Implications for the Use of Meta-Analyses

Christopher J. Ferguson
Texas A&M International University

Michael T. Brannick
University of South Florida

The issue of publication bias in psychological science is one that has remained difficult to address despite decades of discussion and debate. The current article examines a sample of 91 recent meta-analyses published in American Psychological Association and Association for Psychological Science journals and the methods used in these analyses to identify and control for publication bias. Of the 91 studies analyzed, 64 (70%) made some effort to analyze publication bias, and 26 (41%) reported finding evidence of bias. Approaches to controlling publication bias were heterogeneous among studies. Of these studies, 57 (63%) attempted to find unpublished studies to control for publication bias. Nonetheless, those studies that included unpublished studies were just as likely to find evidence for publication bias as those that did not. Furthermore, authors of meta-analyses themselves were overrepresented in unpublished studies acquired, as compared with published studies, suggesting that searches for unpublished studies may increase rather than decrease some sources of bias. A subset of 48 meta-analyses for which study sample sizes and effect sizes were available was further analyzed with a conservative and newly developed tandem procedure of assessing publication bias. Results indicated that publication bias was worrisome in about 25% of meta-analyses. Meta-analyses that included unpublished studies were more likely to show bias than those that did not, likely due to selection bias in unpublished literature searches. Sources of publication bias and implications for the use of meta-analysis are discussed.

Keywords: meta-analysis, publication bias, statistics

Supplemental materials: http://dx.doi.org/10.1037/a0024445.supp

Recent attention to selective publication of positive outcome clinical trials for the efficacy of anti-depressants has highlighted the degree to which publication bias can distort the perceptions of both the scientific community and the public. Several scholarly reviews have revealed that antidepressant trials that show positive effects are published in far greater numbers than are negative or null trials and that some trials with inconclusive findings are discussed in published articles as if clearly positive (Robinson, 2008; Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008). The case of anti-depressants may appear to be easily explained by the potential corrupting influence of economic incentives related to the pharmaceutical industry on the published science. Yet the influence of publication bias on a broader array of fields may distort scientific and public knowledge about topics related to health and many other topics of scientific interest (Rothstein, Sutton, & Borenstein, 2005).

Publication bias refers to the selective publication of studies with a particular outcome, usually those which are statistically significant, at the expense of null studies. A related issue, selection bias, or the tendency of meta-analytic authors to select certain types of studies and not select others for inclusion in meta-analysis, whether intentionally or not, will also be discussed. Selection bias and publication bias do not necessarily co-occur, and it should not be assumed that selection bias implies the presence of publication bias.

It is important to note upfront that not all unpublished studies are unpublished due to obtainment of null results or theory-contradictory findings. Some studies are not published due to major methodological or measurement flaws in the data. Including such deeply flawed studies in meta-analysis is inadvisable as their inclusion would bias the results. Meta-analytic scholars may address this issue by developing and including in the publication clear quality-control criteria for inclusion of unpublished (and published) studies, so long as these criteria are not developed with the intention of giving an advantage to certain outcomes over others.

The issue of publication bias has been identified as important for psychological science for at least three decades (e.g. Rosenthal, 1979). Yet relatively little information is available regarding the extent and impact of publication bias on psychological science, and considerable debate remains about the appropriate methods for psychological science to reduce the impact of publication bias. This article attempts to address some of these gaps in current knowledge by examining publication bias in a sample of 91 recently published meta-analyses. We will consider common approaches to controlling publication bias, namely including unpublished studies and using statistical tests to examine and control for publication bias.

The Use of Unpublished Studies in Meta-Analyses to Control Publication Bias

Published studies in a research field that are not representative of the actual population of all data can result in a biased impression of the research field both in narrative and meta-analytic reviews. Problems with publication bias in pharmaceutical trials have led to calls for researchers to register all randomized clinical trials prior to conducting those trials (Lexchin, Bero, Djulbejovic, & Clark, 2003; Mathieu, Boutron, Moher, Altman, & Ravaud, 2009). However, for much of the basic (and even applied) research in the psychological sciences, implementation of a similar system seems unfeasible, at least in the near future.

One common method of correcting for publication bias in meta-analytic reviews is to conduct a search not only for published studies but for unpublished studies as well. Not surprisingly, many scholars endorse the use of unpublished studies as a correction for publication bias in meta-analytic reviews (e.g. Davis & Crombie, 2001; McLeod & Weisz, 2004; White, 1994; Wilson, 2009). If published studies represent a biased sample of all available studies and if published studies are prone to overestimating effects, as often appears to be the case (e.g. McLeod & Weisz, 2004; Turner et al., 2008), including unpublished studies should do much to correct for the file drawer problem.

However, relatively little data exist regarding the actual effectiveness of this strategy in reducing publication bias effects in meta-analytic reviews. Again, the issue is less whether the *ideal* search for unpublished studies (which in an ideal world would get 100% of them) is effective, and more whether *actual* searches for unpublished studies have been successful in eliminating bias. Indeed, a number of scholars have expressed skepticism or concern regarding whether the inclusion of unpublished studies increases or decreases bias (e.g. Baumeister, DeWall, & Vohs, 2009; Cook et al., 1993; Egger & Davey-Smith, 1998). For instance, although Cook et al. (1993) suggested that inclusion of unpublished studies in meta-analytic reviews can be helpful, the authors noted this is dependent upon finding a full or representative sample of such studies; otherwise, the search may increase rather than decrease bias. There are several issues of potential concern that a search for unpublished studies may not solve.

1. It is possible that unpublished studies may differ from published studies in important methodological ways. For instance, it may be argued that unpublished studies have been declined publication due to significant methodological weaknesses. Under such conditions, including these studies in meta-analyses may be unwarranted and bias the results. One option to address this problem may be quality weighting of studies (Rosenthal, 1991), although it is possible that the criteria selected for "quality" by the meta-analytic authors themselves may merely add a new source of bias.

2. Meta-analytic authors may overrepresent themselves and their close colleagues in a search for unpublished studies. Naturally, locating one's own unpublished studies is far easier than locating them from other sources. Close colleagues may be more inclined to reply to an e-mail or letter than noncolleagues. Even publishing a request to a listserv in the field of specialty may result in a meta-analytic author's receiving papers from a rather homogeneous rather than heterogeneous group of researchers. Thus, even a good faith effort to locate unpublished studies may actually result in a sample of studies that is *more* rather than *less* biased than the population of published studies.

3. Searches for unpublished studies may favor established rather than nonestablished authors. Indeed, it is common for meta-analytic scholars to directly contact prolific scholars in a given field. It would be surprising if this strategy did not result in prolific scholars being heavily represented in the search for unpublished studies. It is possible that doing so may result in a selection of unpublished studies weighted toward the status quo and be less likely to find studies that may challenge the status quo.

4. The search for unpublished studies assumes researcher compliance. Researchers may be noncompliant with requests for unpublished studies for a variety of reasons. They may simply not reply to e-mails or posted requests out of disinterest or time constraints. They may delete requests from unfamiliar scholars without reading them. Unpublished studies may have already been discarded, deleted, or lost. In some cases, researchers may selectively send in only unpublished studies that support their views and not those unsupportive of their views.

Although it is certainly the case that obtaining a full or even representative sample of unpublished studies is of value to meta-analyses (Cook et al., 1993), the lack of a central repository for unpublished studies arguably makes any search for unpublished studies far more problematic than many scholars acknowledge. Indeed, it is argued here that the potential for bias in searches for unpublished studies is high and that, in many cases, the search for unpublished studies may contribute more of a rhetorical fig leaf for the meta-analysis in question than any actual fix for publication bias. This is not to say that excellent and thorough searches for unpublished studies never exist in meta-analytic reviews, as some exemplary examples do exist (e.g. Frattaroli, 2006; Hanson & Morton-Bourgon; 2005; Hawkins, Blanchard, Baldwin, & Fawcett, 2008). However, the degree to which including unpublished studies in meta-analysis is typically effective in reducing bias is a question that could be, but has not yet been, addressed empirically.

A Tandem Procedure for Reducing Type I Error in Publication Bias Analyses

One possibility for reducing Type I error in publication bias analyses is to use a combined tandem procedure between measures. With a tandem procedure, agreement between outcomes enhances confidence in conclusion regarding publication bias. The following tandem criteria were used:

1. Orwin's fail-safe N (FSN) number is lower than k (that is, the number of file-drawer studies needed to reduce the summary effect to trivial size is less than the number of studies in the meta-analysis).

2. Either the rank order correlation or Egger's regression demonstrates significant results.[1]

3. The trim-and-fill method finds and adjusts for missing studies, suggesting evidence of bias. Put succinctly, trim and fill can be used to examine the scatterplot between the sample size and effect size of studies included in a meta-analysis (the "funnel plot"). Symmetry is expected in the plot in the absence of publication bias. The trim-and-fill method imputes effect sizes if the plot is asymmetric. The imputed values are included in a revised overall effect to correct for suspected publication bias. Results that are counterintuitive—that is, those opposing the expected direction of publication bias—should not be considered as evidence for publication bias.

Different approaches to examining publication bias may make Type I errors under different circumstances (Kromrey & Rendina-Gobioff, 2006; Macaskill, Walter, & Irwig, 2001). With a tandem procedure, requiring agreement between disparate indices of publication bias cuts down on the potential rate of Type I error (although Type I error is not eliminated in any analysis, of course). However, using a conservative tandem procedure may increase Type II error rates. The approach is used here to provide a conservative estimate of the relative frequency of encountering publication bias in meta-analyses of interest to psychologists. By examining whether the authors found any evidence of publication bias on the liberal end (more Type I and fewer Type II errors) and using the tandem procedure on the conservative end (fewer Type I and more Type II errors), we created a bracket analogous to a confidence interval for estimating the incidence of publication bias in state-of-the-art meta-analyses in psychology.

To provide an initial test of the operating characteristics of the tandem procedure, we performed a Monte Carlo simulation using the methods described by Kromrey and Redina-Gobioff (2006). The SAS program used by Kromrey and Redina-Gobioff (2006) and provided to us by J. Kromrey (personal communication, December 2010) was modified to include the tandem criteria. The program selects studies by examining p values (significance values for the null hypothesis tests in the primary studies). The program was run with numbers of studies and summary effect sizes similar to those of the sample of 91 meta-analyses described in this article and with no bias (to check Type I error rates) as well as moderate bias (to check power) as documented in Kromrey and Redina-Gobioff (2006). In the no-bias condition, all studies were selected for inclusion in the meta-analysis; for the moderate bias, the function $exp(-2p^{**}1.5)$ was used to generate a probability of inclusion in the meta-analysis. Studies with small p values are very likely to be included, and studies with larger p values have increasingly small chances to be included. Results indicated very low Type I error rates for the tandem procedure (at or below 1% under all conditions).

However Kromrey and Redina-Gobioff reported low power for the trim-and-fill procedure, and similarly low power for the tandem procedure was found using their program, particularly when observed effect sizes were much higher than a critical "trivial" value (i.e., $r = .1$). However other authors have been more sanguine about the performance of trim and fill (e.g., Peters, Sutton, Jones, Abrams, & Rushton, 2006). It is possible that

differences in simulation techniques are part of the explanation of different findings. Peters et al. (2006) noted that trim and fill was designed to find studies deleted because of low effect sizes rather than nonsignificant results, but the Kromrey program simulates bias by choice through significance rather than effect size. Also, the Kromrey program simulates homogeneous underlying (fixed-effects) data, and the trim-and-fill method tends to have greater Type I error rates and better power when effect sizes are heterogeneous (random effects). Further simulation work could be used to investigate such differences, but such work is beyond the scope of this article.

On logical grounds alone, we may conclude that the tandem procedure is conservative. That is, it requires agreement among different indices of bias to conclude that bias is worrisome. The Monte Carlos study provides some empirical support that the approach appears conservative in the ballpark provided by the current sample of studies. Therefore, it seems reasonable that the tandem procedure can provide a lower bound or a conservative estimate of the presence of publication bias.

Publication Bias in a Sample of High-Quality Meta-Analyses

Little concrete information is known about how meta-analytic authors address the issue of publication bias in psychological science. Similarly, the extent of publication bias in psychological science remains speculative. To provide data in response to these issues, the current study sampled recent meta-analyses from top journals in psychology. Data analyzed in this sample of meta-analyses will help answer how publication bias is being addressed in psychological science, whether inclusion of unpublished studies "fixes" publication bias, and whether unpublished studies are more or less extreme than published studies, and will provide some estimates of the prevalence of publication bias. Our current analyses attempt to address the following questions:

1. Are authors attending to publication bias?

2. Does publication bias appear to be a problem in meta-analyses in psychology?

3. How do authors combat publication bias?

Method

Selection of Meta-Analyses

We searched the PSYCINFO data base in November of 2009 using an abstract search for the term *meta**. The 5-year time period of November 2004–November 2009 was used to get a sample of recent meta-analyses, as it was assumed that more recent meta-analyses would reflect current state of the art. To that end, we limited the search for meta-analyses to several top American Psychological Association (*American Psychologist*, *Developmental Psychology*, *Journal of Abnormal Psychology*, *Journal of Con-*

---

[1] Requiring both of these to be statistically significant would further tighten the concordance, but given the low power of both these tests, many true publication bias scenarios would be missed. Thus, this is not recommended.

sulting and Clinical Psychology, *Journal of Personality and Social Psychology*, *Psychological Bulletin*) and Association of Psychological Science (*Perspectives on Psychological Science*, *Psychological Science*) journals. All of the meta-analyses that met the previously described criteria were included. The intent was to collect a representative sample of these state-of-the-art meta-analyses, not necessarily every meta-analysis published during this time frame. Given that these meta-analyses are published in the most selective and well-respected journals in psychology, we felt this sample would reflect "best practices" approaches in psychological meta-analytic research and would allow the current analysis to answer the question: What are current best practices in psychological meta-analyses, and how effective are these in addressing the issue of publication bias? A sample of 91 meta-analyses meeting the inclusion criteria was ultimately retrieved and these are listed in.

Study Coding

From each study, several pieces of information were coded: number of total manuscripts included in the meta-analysis, number of dissertations, number of other unpublished sources (including unpublished data, conference presentations, master's theses, in-press manuscripts), whether the authors had searched for publication bias using any techniques, and whether the authors reported any evidence suggestive of publication bias. Regarding unpublished manuscripts, we treated dissertations as a distinct category as many (but not all) dissertations are locatable through the Dissertation Abstracts database and thus are indexed and retrievable in a nonbiased search. By contrast, other unpublished works (including master's theses and in-press manuscripts) are not able to be searched through a central abstracting repository. Although it naturally may be argued that in-press manuscripts are more similar in quality to published works, as in-press manuscripts are not yet indexed, they are likely to be as difficult to locate as other unpublished works and were thus included in this category. In other words, the hypothesis being tested here was whether searches for nonindexed manuscripts increased bias in meta-analyses. This should not be interpreted as an assumption that in-press and unpublished works are similar other than in their status as being nonindexed at the time of the meta-analytic search. One question to be addressed was whether nonindexed nonabstracted works had greater or less bias than published works. The numbers of published and unpublished works authored by the meta-analytic authors themselves were ascertained. Once these data were collected, the authors of the meta-analyses were contacted with these figures to identify and correct any potential discrepancies.

From these data were calculated the proportion of published articles authored by the meta-analytic authors themselves, the proportion of unpublished manuscripts published by the authors themselves, and the proportion of unpublished studies (including both dissertations and other unpublished works) to total number of studies included in each meta-analysis.

Some of the meta-analyses included in the sample described earlier included individual study sample sizes and effect sizes necessary for a basic recreation of their meta-analyses. We augmented this number by contacting the meta-analytic authors and requesting these data when it was not available in the original manuscript. Data on 48 meta-analyses were thus available for

further analyses. These meta-analyses were recreated from data supplied using the <cr4>Comprehensive Meta-Analysis program (Version 2). These meta-analyses were tested with the tandem procedure described previously, allowing for a more consistent analysis of publication bias across meta-analyses. In some manuscripts, multiple related meta-analyses were reported together. For instance, in a single manuscript, meta-analyses may have been computed and reported separately across sample characteristics or differing outcomes. In each of these manuscripts, the analysis with the largest number of individual samples was selected for recreation here.

Results

Are Meta-Analytic Authors Attending to Publication Bias, and How Do They Address It?

Of the 91 studies included in the current analysis, 64 (70%) showed that there was at least some effort to evaluate publication bias. Methods employed were highly heterogeneous, however, ranging from mean comparisons of unpublished to published studies, visual examination of the funnel plot (a visual display of the relationship between sample size and effect size), to use of FSN through use of statistical funnel plot analyses. In a much smaller number, 28 (31%), more than one analytic strategy was employed. Table 1 presents an analysis of how often specific means of attending to publication bias were used.

Of those studies in which any analysis of publication bias was conducted, 26 (41%) reported finding some evidence of publication bias.

The Use of Unpublished Studies

Of the meta-analyses included in the current review, 49 or just over half (54%) employed doctoral dissertations as a source of unpublished studies. A smaller number of meta-analyses (41, or 45%) included unpublished studies from nonindexed sources. A majority of studies (57, or 63%) included unpublished studies from

Table 1

Publication Bias Procedures Used in Included Meta-Analyses and the Proportion of Studies Using Them That Found Evidence of Bias

Procedure used	Proportion of meta-analyses	Positive hit rate for technique
Comparison of published/unpublished studies effect sizes	19 (21%)	35%
Visual examination of funnel plot	5 (5%)	0%
Fail-safe *N*	20 (22%)	35%
Orwin's fail-safe *N*	7 (8%)	14%
Begg and Mazumdar's (1994) rank-correlation	4 (4%)	50%
Egger's regression	5 (5%)	50%
Trim and fill	22 (24%)	64%
Other, less broadly used procedures	13 (14%)	15%
No publication bias procedure	27 (30%)	N/A
One procedure	38 (42%)	53%
Two procedures	23 (25%)	30%
More than two procedures	3 (3%)	0%

one source or another (either doctoral dissertations or nonindexed sources).

However, for those meta-analyses that did include unpublished studies, the number of unpublished studies included was often small. As can be seen in Figure 1, the largest number of meta-analyses had unpublished studies as no more than 5% of the total number of studies. By and large, the distribution for the proportion of unpublished studies is positively skewed with very few meta-analyses approaching an equal number of published and unpublished studies. It is worth noting that it is not necessarily the case that an equal number of published and unpublished studies exist. The actual number of unpublished studies may be greater than, roughly equal to, or less than published studies and may differ markedly across research fields.

Regarding whether samples of nonindexed unpublished studies (that is, unpublished studies excluding dissertations) reduce or increase bias in meta-analyses, one source of potential bias is the proportion of unpublished studies authored by the meta-analytic scholars themselves. In this case, doctoral dissertations were not considered, as these are indexed in Dissertation Abstracts and thus less susceptible to selection bias than other sources of unpublished studies. Although it is not possible to quantify manuscripts from close associates through available electronic databases, it is possible to quantify manuscripts authored by the meta-analytic scholars themselves. In the sample of 91 meta-analyses described here, on average, meta-analytic scholars were represented in 5.89% of published works included in their meta-analysis. By contrast, the representation of meta-analytic scholars in nonindexed unpublished works included in their meta-analyses was more than double this amount at 12.94%.

Regarding whether meta-analyses that included unpublished studies (including doctoral dissertations and other forms of unpublished studies) were more or less likely to report finding evidence of publication bias, we performed a simple phi-coefficient correlation (use of unpublished studies by bias outcome). The resultant correlation was nonsignificant and near zero ($r = -.02$), suggesting that the inclusion of unpublished studies did not influence a finding of publication bias. A point-biserial correlation between the proportion of unpublished studies and findings of publication bias was similarly nonsignificant ($r = -.06$). When we performed the phi-coefficient correlation again with dissertations removed from the category of unpublished studies, the relationship between inclusion of nonindexed unpublished studies and reported presence of publication bias remained nonsignificant ($r = .08$).

Is There a Problem? A Subset of Analyses Using the Tandem Procedure

We reanalyzed and assessed a subset of 48 included meta-analyses using the tandem procedure described earlier in the article (both random and fixed effects were calculated, although random effects were used consistently for comparison purposes). As indicated earlier, these 48 meta-analyses were all those that reported effect size data on individual studies included in the meta-analysis and additionally those studies for which the authors responded to a request for such data. The remaining studies did not report these data, and the authors did not respond to (or declined) requests for these data. Naturally it is possible that this subset of 48 meta-analyses may differ from the 43 for which corresponding data were not available. However, neither the average number of studies (71.3 vs. 69.7) nor the average reported effect size ($r = .19$ vs. .21) differed significantly (t test values of 0.08 and -0.47, respectively, $p > .05$ for both), suggesting an absence of systematic difference between the 48 reanalyzed studies and the 43 for which data were not available.

Using the tandem procedure on the subsample of 48 meta-analyses for which individual study data were available suggested that 12 (25%) experienced potential publication bias. Thus, use of the tandem procedure appears to be considerably more conservative in the identification of publication bias than the heterogeneous methods employed by authors themselves.

Using results from the tandem procedure, we again calculated a phi-coefficient correlation between the use of unpublished studies and presence of publication bias. Results indicated that meta-analyses that included unpublished studies actually were more likely to demonstrate publication bias ($r = .27$, $p < .05$) than were meta-analyses that did not include unpublished studies. Removing dissertations and considering only nonindexed unpublished studies
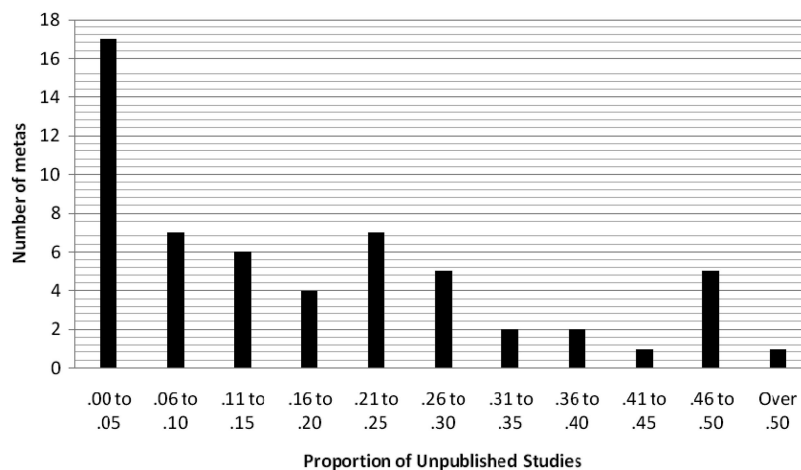

Figure 1.   Proportions of unpublished studies included in meta-analyses.

produced similar results ($r = .31$, $p < .05$). Similar results were found when the proportion of unpublished studies were considered ($r = .30$, $p < .05$) rather than a dichotomous variable. When only those meta-analyses that included unpublished studies were considered, the correlation between proportion of unpublished studies and evidence for publication bias was .2 ($p < .05$). Taken together, these results suggest that including unpublished studies in meta-analyses is potentially problematic, at least under current practice. Of course, the inclusion of unpublished studies cannot properly be called *publication bias* but represents a more general *selection bias* in the meta-analysis itself. Put more simply, evidence presented here suggests that searches for unpublished studies increase findings of selection bias, likely due to the difficulty in obtaining a representative sample of unpublished studies.

We took a closer look at the dissertations and unpublished articles included in these meta-analyses. The average effect size, across meta analyses, for published articles was approximately $r = .20$ ($SD = .12$). For dissertations, this was lower at $r = .13$ ($SD = .16$), and for unpublished articles, $r = .17$ ($SD = .18$). Unpublished studies appeared more similar in effect size to published studies than did dissertations, possibly because authors other than doctoral students will not trouble to write up a study they perceive to be difficult to publish. A dependent t test for the difference in effect size between published studies and dissertations was significant, $t(24) = 2.74$, $p = .012$, $r = .26$, 95% confidence interval (CI) $= .07 < r < .43$. The difference between published and unpublished studies was nonsignificant, $t(15) = 1.04$, $p = .32$, $r = .13$, 95% CI $= -.14 < r < .37$. These effect-size differences are arguably small, suggesting that for meta-analyses with large effect size results, the inclusion or noninclusion of a robust sample of unpublished studies may have little bearing on conclusions. However, for meta-analyses at the threshold of "trivial" effects (which in fact corresponded to the majority of meta-analyses), the difference could have a substantial effect on conclusions.

Figure 2 plots the relationship between published studies and dissertations in the included meta-analyses. Were dissertations to come from the same population of effect sizes as published studies, there should be an equal distribution of plotted points above and below the line, but this is not the case. Figure 3 documents the same data for unpublished studies. These results indicate that published studies are not representative of the total population of effect sizes, although dissertations may be a better representation of the missing studies than are nonindexed unpublished studies.

Discussion

Results presented as part of this article form, to the best of our knowledge, the first assessment of publication bias in meta-analyses published in leading scientific journals in psychology. We are able to describe evidence regarding the prevalence of worrisome levels of publication bias and authors' customary methods of dealing with it.

Are Authors Attending to Publication Bias?

The majority of meta-analytic authors appeared to attend to the issue of publication bias, but nearly a third did not. Despite a general awareness of the issue, efforts to identify and combat publication bias currently employed are highly heterogeneous;
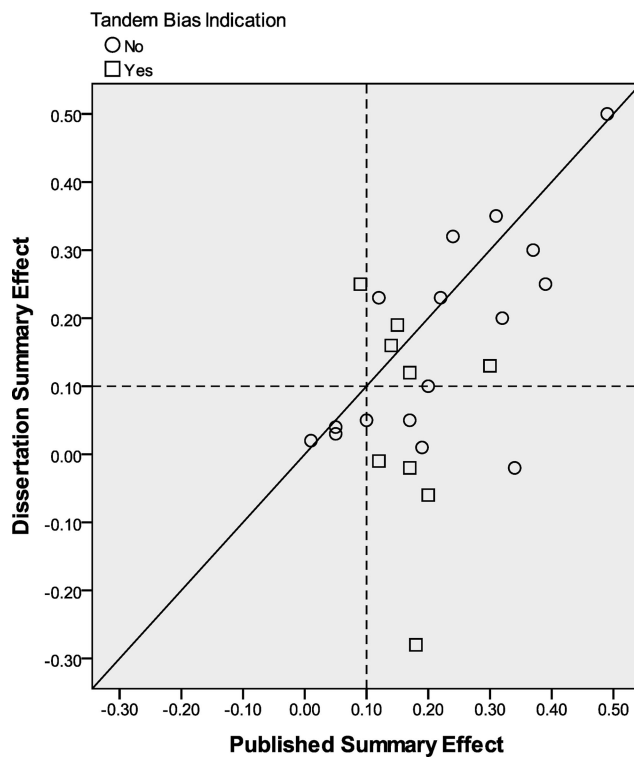

Figure 2. Plot of published and dissertation effect sizes in individual meta-analyses. Tandem bias indication = whether the tandem procedure indicated the specific meta-analyses showed publication bias.

authors rarely choose more than one approach, and the rationale for the choice is rarely made clear. Different methods of assessing publication bias need not agree with each other; some clearly have different meanings than do others. For example, the trim-and-fill method judges the symmetry of the distribution of effect sizes and imputes values on the low end, which results in a reduced summary effect. The Orwin's FSN method, on the other hand, does not assess the distribution for bias at all but instead starts from the assumption that all the available studies are erroneous and calculates how many missing studies would be needed to reduce the summary effect to a given level. Thus, meta-analysts (and readers) should not assume that the results of one test will tell much about the results of another. Because it is impossible to know the values of effect sizes that are missing, it is reasonable to conduct sensitivity analyses to consider how fragile the summary effect might be. There are sophisticated methods available for such analyses (e.g., Vevea & Woods, 2005), but the authors of the meta-analyses in our sample rarely used them (the rare exceptions were Pettigrew & Tropp, 2006; Shadish & Baldwin, 2005).

The results from individual publication bias tests appear to overlap imperfectly. Considering a single analysis alone may result in erroneous conclusions. Bias can result from more processes than the selective publication of significant results, including general ideological biases within a research field, the influences of a single or collaborating group of research labs that are highly productive but do not necessarily represent the broader field, choices made by researchers on how they analyze data in individual studies, and so forth.
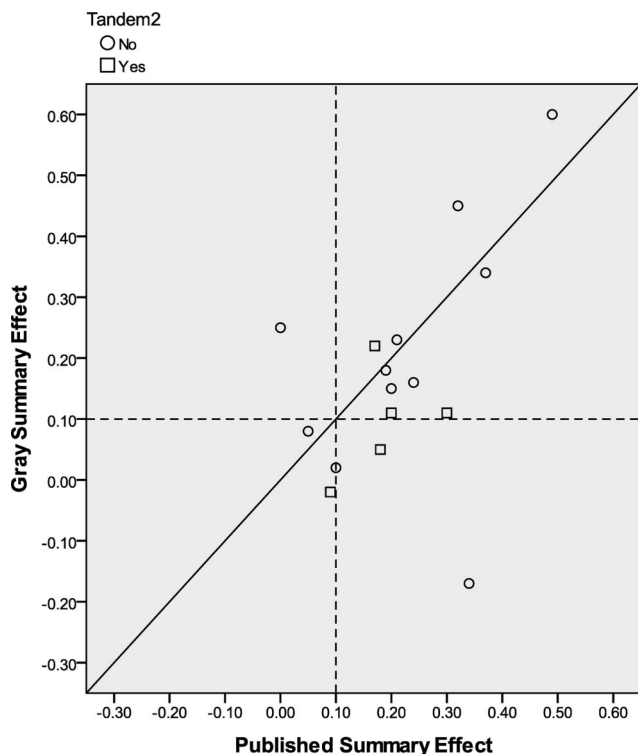
Figure 3. Plot of published and unpublished (gray) effect sizes in individual meta-analyses. Tandem bias indication = whether the tandem procedure indicated the specific meta-analyses showed publication bias.

Does Publication Bias Appear to Be a Problem?

Yes, depending on the method of estimation, approximately from 20% to 40% of the meta-analyses reported a publication bias issue. The low-end figure came from the tandem procedure, in which a study is not flagged unless there are multiple indications of publication bias; thus, many meta-analyses in psychology exhibit both asymmetric funnel plots and small enough summary effects to appear fragile in the presence of publication bias. Thus, there is reason for meta-analysts to continue to investigate and report issues of publication bias in their studies. As noted by Rothstein et al. (2005), there are additional reasons for authors to consider publication bias routinely, including reader skepticism about the credibility of the summary result of a meta-analysis.

How Do Authors Combat Publication Bias?

About 60% of the authors searched for unpublished studies to include in their analysis. In a situation where all research trials must be registered in a central index prior to the study implementation, inclusion of unpublished works is least likely to be biased, since all unpublished works are indexed and known. In the absence of a central registry, obtaining an appropriate and nonbiased sample of unpublished studies may be difficult, even when authors work very hard to do so. Meta-analytic authors who do not have the benefit of a central registry will need to be especially diligent in their searches for unpublished studies. Results from the current study suggest that the majority of such searches are rather rudi-

mentary at best. Few meta-analyses find proportions of unpublished studies that equal published studies in number. In the current analysis, the greatest number (i.e., mode) of meta-analyses included unpublished studies at 5% or less of the total included studies. Although the true number of unpublished studies is unknowable, it is difficult to assert in the competitive world of academic publishing that the majority of studies are published and that the majority of these happily happen to be statistically significant. Thus, we raise the concern that some searches for unpublished studies are ineffective and unintentionally biased, although this observation must be qualified with the observation that many unpublished studies are themselves inherently flawed and, thus, should not be included in meta-analyses.

Further research may help elucidate means for improving the selection of unpublished studies. One option may be for meta-analytic authors to recruit or hire independent searchers or search professionals to conduct thorough literature searches on their behalf (see Campbell Collaboration, 2010, as an example). It may further be incumbent upon journals and peer reviewers to pay greater attention to the results of searches for unpublished studies so as to assess whether such searches appear to be successful. It may be simply that too often that the meta-analytic authors' claims to have searched for unpublished studies mollify reviewers, who do not attend closely to the results of such searches. In fairness, we did not assess in this article whether the literature searches by authors of meta-analyses are more or less rigorous than the literature searches conducted by other authors, such as those of narrative reviews. The problems identified here may extend more generally to narrative reviews as well.

It may behoove meta-analytic authors to pay greater attention to the distinction between cursory and exhaustive searches for unpublished studies. A cursory search, turning up very few unpublished studies (perhaps only a tiny fraction of those that exist), may do little to address the inequities in the published studies while presenting a veneer of having done so. Indeed, a cursory search for unpublished studies may turn up a sample of such studies that is highly selective and arguably as biased as the included published studies. More rigorous and exhaustive searches may result in fewer problems in this realm. This issue may be particularly important when meta-analytic authors are examining a debated research field, particularly when the authors themselves are on "one side" of the debate. Failing to rigorously secure publications or unpublished studies from both sides of a contentious field may do little to resolve underlying debates and inconsistencies.

Although more rigorous searches are likely better than cursory searches in providing a complete picture of the research literature, researchers often will avoid writing an article at all if they think the chance that it will be published is small. Thus, there may be effect sizes for which there are no studies (other than perhaps a moldy computer printout) lingering in file drawers, and such effect sizes will not likely accrue through even the most thorough searches.

What Is the Effect of Including Unpublished Studies?

Unpublished (gray) literature showed nonsignificantly smaller summary effect sizes on average compared with their published counterparts. Results from the current analysis also suggest that included gray studies tend to suffer from selection bias. Because

they are not indexed, that gray studies should suffer from selection bias should probably not be surprising. Current results suggest that meta-analytic authors are twice as often represented in unpublished studies (aside from dissertations) than published studies. Although difficult to quantify, it is reasonable to assume that this phenomenon extends to close colleagues of the meta-analytic authors who may be both better known and more likely to respond to requests for unpublished papers. Although it is reasonable to suspect that this phenomenon is partly attributable to selection bias, we suggest that is also possible that a meta-analytic author's own work may simply best represent the kind of inclusion criteria that the author sets for the meta-analysis in question. For example, outcome measures, study designs, hypotheses, and so forth that interest a particular meta-analytic author would also likely influence his or her own studies. The current study cannot answer the question of whether studies by meta-analytic authors differ systematically from other studies (particular unknown unpublished studies that may have been missed). It is also unclear the degree to which author-published studies bias meta-analyses in the statistical sense.

Dissertations, being indexed, are less susceptible to selection bias problems. The customary statistical analyses of effect sizes in meta-analysis (e.g., the placement of confidence intervals about the mean effect size) rest on the assumption that the studies in the meta-analysis constitute a random sample of studies from some population (e.g., Bonett, 2008). Although the assumption of random sampling of studies appears difficult to justify, dissertations appear closer to a representative sample of all studies completed about a given research question than do studies published in refereed journals, based on what we know of rejection rates for dissertations and journal articles. Note that in Figure 2, summary effects based on published articles are more likely to fall above the .10 (dashed line) threshold for meaningfully large effects than are summary effects from dissertations (80% vs. 56%). The average difference in summary effects between published studies and dissertations was .08 correlation units in this sample. Such a difference is large enough to matter when the summary effect is near the border of meaningfulness, which is often the case in meta-analyses in psychology.

On the other hand, it could be argued that studies in journals are more representative of the studies of interest than are the dissertations because the published studies are of better quality (consider the arguments for best evidence synthesis; Slavin, 1986). For example, one might consider random assignment to treatment as an inclusion criterion, so that the question becomes how well studies in the meta-analysis represent the larger population of studies in which random assignment to treatment is used.

The position taken in this article is that differences between publications and dissertations are reasonably indicative of nonrepresentative sampling because the meta-analytic authors' own inclusion criteria were used in both cases. That is, Figure 2 shows comparison between dissertations and published studies selected with the same inclusion criteria. Thus, if the design was considered by the author as suitable for inclusion in published articles, then it is also suitable for inclusion as a dissertation, and differences in effect sizes could reasonably be considered as indicative of publication bias, although we cannot rule out other potential explanations of the difference.

Alternatives to Publication Bias

It remains possible that publication bias analyses do not adequately distinguish "true" publication bias from "small-study" effects. In the current study, we did not specifically examine small-study effects (effects due to systematic methodological differences between smaller N and larger N studies). However, positive findings (i.e., an asymmetrical relationship between sample size and effect size) due to small-study effects can be just as detrimental to a meta-analysis as publication bias. Small-study effects can be examined through careful analysis by meta-analytic authors of the studies included in the meta-analysis. Should small and large studies be found to truly differ on substantial methodological issues, it may be best for the meta-analysis to be broken in two across the methodological divide (not necessarily by sample size) and reassessed for publication bias. The meta-analytic author who includes a heterogeneous sample of methodologically diverse studies into a meta-analysis and then dismisses positive publication bias findings as small-study effects is being somewhat disingenuous. It is imperative for meta-analytic authors to examine, explain, and control for small-study effects if they believe they exist in their meta-analyses. Other issues that may bias findings, such as membership of study authors in particular theoretical camps of a controversial field, may also be worth considering as moderator variables.

Concluding Statements

The current review examined the issue of publication bias in psychological science. Efforts to identify and control publication bias in current research remain heterogeneous and, at times, perfunctory. The common practice of including unpublished studies as a control for publication bias appears to be ineffective and, at times, counterproductive, at least in current practice. Although some researchers have indeed been diligent and apprehensive about the potential influence of bias in science, this review concludes with the concern that psychological science has not yet tackled this issue in full. To some degree, it may not seem in psychological science's best interest to address publication bias directly, as the results include reduced effect size estimates and open identification of larger issues of subjectivity and bias in psychological science that are painful to confront and potentially harmful to the reputation of psychology. However, at its fundamental root, psychological science, like any science, must concern itself with the truth. In tackling the issue of bias in our field, we come one step closer.

References

Baumeister, R., DeWall, C., & Vohs, K. (2009). Social rejection, control, numbness, and emotion: How not to be fooled by Gerber and Wheeler (2009). *Perspectives on Psychological Science, 4,* 489–493. doi:10.1111/j.1745-6924.2009.01159.x

Bonett, D. G. (2008). Meta-analytic interval estimation for bivariate correlations. *Psychological Methods, 13,* 173–181. doi:10.1037/a0012868

Campbell Collaboration. (2010). *What is a systematic review?* Retrieved from http://www.campbellcollaboration.org/what_is_a_systematic_review/index.php

Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Cook, D. J., Guyatt, G. H., Ryan, G., Clifton, J., Buckingham, L., Willan, A., . . . Oxman, A. D. (1993). Should unpublished data be included in meta-analyses? Current convictions and controversies. *Journal of the American Medical Association, 269,* 2749–2753.

Davis, H., & Crombie, I. (2001). What are meta-analyses? *What Is . . .,* 1, 1–8. Retrieved from http://meds.queensu.ca/medicine/obgyn/pdf/what_is/WhatisMetaAn.pdf.

Egger, M., & Davey-Smith, G. (1998). Meta-analysis: Bias in location and selection of studies. *British Medical Journal, 316,* 61. Retrieved from http://www.bmj.com/content/316/7124/61.full

*Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis. *Psychological Bulletin, 132*(6), 823–865. doi:10.1037/0033-2909.132.6.823

*Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73,* 1154–1163. doi:10.1037/0022-006X.73.6.1154

*Hawkins, A. J., Blanchard, V., Baldwin, S., & Fawcett, E. (2008). Does marriage and relationship education work? A meta-analytic study. *Journal of Consulting and Clinical Psychology, 76,* 723–734. doi:10.1037/a0012584

Kromrey, J., & Rendina-Gobioff, G. (2006). On knowing what we do not know: An empirical comparison of methods to detect publication bias in meta-analysis. *Educational and Psychological Measurement, 66,* 357–373. doi:10.1177/0013164405278585

Lexchin, J., Bero, L., Djulbejovic, B., & Clark, O. (2003). Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *British Medical Journal. 326.* 1167–1170. doi:10.1136/bmj.326.7400.1167

Macaskill, P., Walter, S., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine, 20,* 641–654. doi:10.1002/sim.698

Mathieu, S., Boutron, I., Moher, D., Altman, D., & Ravaud, P. (2009). Comparison of registered and published primary outcomes in randomized controlled trials. *Journal of the American Medical Association, 302,* 977–984. doi:10.1001/jama.2009.1242

McLeod, B. D., & Weisz, J. (2004). Using dissertations to examine potential bias in child and adolescent clinical trials. *Journal of Consulting and Clinical Psychology, 72,* 235–251. doi:10.1037/0022-006X.72.2.235

Peters, J. L., Sutton, A., Jones, D., Abrams, K., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between study heterogeneity. *Statistics in Medicine, 26,* 4544–4562. doi:10.1002/sim.2889

*Pettigrew, T., & Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90,* 751–783. doi:10.1037/0022-3514.90.5.751

Robinson, D. (2008). Evidence-based medicine and publication bias. *Primary Psychiatry, 15,* 30–33.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641. doi:10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). Quality-weighting of studies in meta-analytic research. *Psychotherapy Research, 1,* 25–28. doi:10.1080/10503309112331334031

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.) (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments.* New York, NY: Wiley. doi:10.1002/0470870168

*Shadish, W., & Baldwin, S. (2005). Effects of behavioral marital therapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology, 73,* 6–14. doi:10.1037/0022-006X.73.1.6

Slavin, R. E. (1986). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15,* 5–11.

Turner, E. H., Matthews, A., Linardatos, E., Tell, R., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine, 358,* 252–260. doi:10.1056/NEJMsa065779

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10,* 428–443. doi:10.1037/1082-989X.10.4.428

White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & Hedges, L. V. (Eds.) *The handbook of research synthesis* (pp. 41–56). New York, NY: Russell Sage Foundation.

Wilson, D. (2009). Missing a critical piece of the pie: Simple document search strategies inadequate for systematic reviews. *Journal of Experimental Criminology, 5,* 429–440. doi:10.1007/s11292-009-9085-5