PERSONNEL PSYCHOLOGY
THE STUDY OF PEOPLE AT WORK

PERSONNEL PSYCHOLOGY
2012, 65, 221–249

REVISITING THE FILE DRAWER PROBLEM IN META-ANALYSIS: AN ASSESSMENT OF PUBLISHED AND NONPUBLISHED CORRELATION MATRICES

DAN R. DALTON
Department of Management and Entrepreneurship
Kelley School of Business
Indiana University

HERMAN AGUINIS
Department of Management and Entrepreneurship
Kelley School of Business
Indiana University

CATHERINE M. DALTON
Department of Management and Entrepreneurship
Kelley School of Business
Indiana University

FRANK A. BOSCO
Department of Management, Marketing, and MIS
Lewis College of Business
Marshall University

CHARLES A. PIERCE
Department of Management
Fogelman College of Business and Economics
University of Memphis

The file drawer problem rests on the assumption that statistically nonsignificant results are less likely to be published in primary-level studies and less likely to be included in meta-analytic reviews, thereby resulting in upwardly biased meta-analytically derived effect sizes. We conducted 5 studies to assess the extent of the file drawer problem in nonexperimental research. In Study 1, we examined 37,970 correlations included in 403 matrices published in *Academy of Management Journal* (*AMJ*), *Journal of Applied Psychology* (*JAP*), and *Personnel Psychology* (*PPsych*) between 1985 and 2009 and found that 46.81% of those correlations are not statistically significant. In Study 2, we examined 6,935 correlations used as input in 51 meta-analyses published in *AMJ*, *JAP*,

A previous version of this paper was presented at the meetings of the Academy of Management, San Antonio, TX, August 2011. Also, a previous and much abbreviated version excluding detailed numerical results reported in this paper was published in the *Academy of Management Best Paper Proceedings* (2011). We thank the generosity of each of the researchers who anonymously and so gracefully provided us with data for Study 3. These individuals serve as an example of scholarship and collegiality.

Correspondence and requests for reprints should be addressed to Herman Aguinis, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Suite 630D, Bloomington, IN 47405–1701; haguinis@indiana.edu.

© 2012 Wiley Periodicals, Inc.

PPsych, and elsewhere between 1982 and 2009 and found that 44.31% of those correlations are not statistically significant. In Study 3, we examined 13,943 correlations reported in 167 matrices in nonpublished manuscripts and found that 45.45% of those correlations are not statistically significant. In Study 4, we examined 20,860 correlations reported in 217 matrices in doctoral dissertations and found that 50.78% of those correlations are not statistically significant. In Study 5, we compared the average magnitude of a sample of 1,002 correlations from Study 1 (published articles) versus 1,224 from Study 4 (dissertations) and found that they were virtually identical (i.e., .2270 and .2279, respectively). In sum, our 5 studies provide consistent empirical evidence that the file drawer problem does not produce an inflation bias and does not pose a serious threat to the validity of meta-analytically derived conclusions as is currently believed.

In one of the seminal treatments of meta-analysis, Hunter, Schmidt, and Jackson (1982) observed that "scientists have known for centuries that a single study will not resolve a major issue. Indeed, a small sample study will not even resolve a minor issue. Thus, the foundation of science is the cumulation of knowledge from the results of many studies" (p. 10). That entreaty for research syntheses, overwhelmingly in the form of meta-analytic approaches, has led to a veritable revolution in methodology and analyses in organizational behavior and human resource management (OBHRM), industrial and organizational (I-O) psychology, and many other fields (e.g., Aguinis, 2001; Burke, 1984; Burke & Landis, 2003; Le, Oh, Shaffer, & Schmidt, 2007; Mount, Oh, & Burns, 2008; Schmidt, Oh, & Hayes, 2009; Schmidt, Shaffer, & Oh, 2008).

If we consider modern meta-analysis to have been initially developed in the late 1970s (Cooper, 1979; Glass, 1977; Schmidt & Hunter, 1977; Smith & Glass, 1977), consider its growth in the 40-year period since. For the period 1980–2010, there are 5,183 articles with the expression "meta-analysis" or its derivatives in the PsycINFO database. The EBSCO Academic/Business Source Premier database notes 10,905 such articles. In addition, the MedLine database features 15,627 such entries. Another indicator of the interest in meta-analyses is the recent proliferation of special issues and compendia on this topic (e.g., Aguinis, Pierce, Bosco, Dalton, & Dalton, 2011; Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, Hedges, & Valentine, 2009; Curran, 2009; Geyskens, Krishnan, Steenkamp, & Cunha, 2009; Rousseau, Manning, & Denyer, 2008). In addition, consider the coverage of meta-analyses in the *Annual Review of Psychology*, one of the most influential publications across all fields in psychology (i.e., the impact factor released in June 2011 was 18.29, which indicates the mean number of citations received during 2010 by articles published in 2008 and 2009). *Annual Review of Psychology* chapters place more emphasis on meta-analytic compared to primary level study results (S. T. Fiske, personal communication, November 1, 2007). Not surprisingly, meta-analyses receive three times as many citations as primary-level studies (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011).

Notably, from the onset of this remarkable growth in the reliance on meta-analyses for research syntheses, there has been an enduring issue that is believed to compromise the fidelity of results generated from these approaches—the file drawer problem (e.g., Greenwald, 1975; Rosenthal, 1979). The file drawer problem rests on the assumption that statistically nonsignificant results are less likely to be published and, hence, less likely to be included in meta-analytic reviews, thereby resulting in an upwardly biased sample of primary-level effect-size estimates and upwardly biased meta-analytically derived summary effect sizes. There is a strong and enduring belief that the file drawer problem is an important cause for concern that compromises meta-analytic results and, hence, substantive conclusions with important consequences for theory and practice. For example, consider the following illustrative statements from articles published in *Journal of Applied Psychology* (*JAP*) and *Personnel Psychology* (*PPsych*). Mone, Mueller, and Mauland (1996) predicted that "statistically underpowered research may never overcome what Rosenthal (1979) described as the 'file drawer problem,' meaning that due to a field's obsession with statistical significance, a study's nonsignificant findings may relegate it to terminal existence in a file drawer" (p. 117). Viswesvaran, Barrick, and Ones (1993) issued the warning that "a practical concern in meta-analysis is that the studies being cumulated may not be representative of all the studies conducted examining that relationship. Rosenthal (1979) and Orwin (1983) provide expressions for estimating the number of null or nonsignificant studies that would have to be withheld in file drawers of researchers (i.e., usually the unpublished primary studies) to threaten the conclusions derived from a meta-analytic cumulation. The meta-analyst is then required to judge whether the conclusions of the meta-analysis are likely to be reversed if all the data were available" (pp. 555–556). Similarly, Gilboa, Shirom, Fried, and Cooper (2008) cautioned that "several researchers have raised the possibility that meta-analytic studies may produce inaccurate (e.g., upwardly biased) estimates of the relationships in question because of the publication bias, which reflects the premise that studies producing non-significant or unexpected results are less likely to be submitted and less likely to be accepted for publication" (p. 246). In addition to the importance of the issue in OBHRM, I-O psychology, and related fields, the file drawer problem is considered to be a very serious threat in many other scientific fields (e.g., neuroscience). As a recent illustration, Fiedler (2011) issued the warning that "a file-drawer bias (Rosenthal, 1979) facilitates the selective publication of strong correlations while reducing the visibility of weak research outcomes" (p. 167). Moreover, Fiedler (2011) argued that "[w]hat we are dealing with here is a general methodological problem that has intrigued critical scientists under many different labels: the file-drawer bias in publication (Rosenthal, 1979)" (p. 164).

These and numerous other similar statements that are included routinely in published meta-analyses have two characteristics in common. First, they argue that the file drawer problem is pervasive and almost unavoidable. In other words, the source of the file drawer problem is in the effect-size estimates reported in primary-level research. Because meta-analysts do not have access to the original data but only to the supposedly upwardly biased resulting effect-size estimates, the file drawer is an insurmountable problem. Second, the file drawer problem is viewed as a critical issue to consider when conducting a meta-analysis because it has an important biasing effect on the resulting effect-size estimates. This bias in the resulting meta-analytic results is an important problem for theory and practice. Specifically, biased effect-size estimates lead to theory derailments and practices that may not be as effective as expected. For example, practitioners may implement selection, training, and other interventions incorrectly believing that the effectiveness of such interventions will be greater than they actually are due to an assumed overestimation of meta-analytically derived effect sizes. This overestimation and incorrect belief are supposedly caused by the file drawer problem. In short, the file drawer is considered to be a pervasive and important problem.

What if it were possible to conduct a study to test the extent to which the file drawer problem exists and biases meta-analytic results? Moreover, what if the file drawer problem does not actually exist? What if, contrary to the field's *zeitgeist*, the file drawer problem is a methodological myth and urban legend as has been suggested regarding other established methodological practices (Lance, 2011)? If the file drawer is not the big problem it is believed to be, there would be important implications for research and practice. With respect to research, future meta-analyses would not need to implement procedures to estimate the extent to which the file drawer may have biased the resulting effect-size estimates. As we describe in subsequent sections, concern for the biasing effects of the file drawer problem has led to a vast literature on how to mitigate its effects, and meta-analysts routinely implement such procedures. If the file drawer is not really the problem assumed to be, we would have more confidence that meta-analytically derived effect sizes are not upwardly biased. As we also describe later in this paper, the fear of the biasing effects of the file drawer problem has led meta-analysts to include caveats, warnings, and cautionary statements regarding practices based on meta-analytic results given the potential overestimation of effects.

Next, we provide a brief synopsis of the file drawer problem and describe a five study research program that offers a novel approach to revisiting the long-held belief that the file drawer problem causes an upward bias in meta-analytically derived effect sizes. Indeed, via collecting data from six different sources, and engaging in methodological triangulation, we find that, contrary to the established belief, the file drawer problem is of little, if any, consequence for meta-analytically derived theoretical conclusions and applications in OBHRM, I-O psychology, and related fields.

File Drawer Problem

The file drawer problem is a subset of a broader category of publication biases. The issue of publication bias potentially arises "whenever the research that appears in the published literature is systematically unrepresentative of the population of completed studies" (Rothstein, Sutton, & Borenstein, 2005, p. 1; see Dickerson, 2005, and Sutton, 2005, 2009 for extended histories of publication bias). This is not simply a missing-at-random data problem. That a given study was randomly excluded from a meta-analysis would be of little consequence. In such a case, the synthesis would be relying on a sample of the available research/data. As noted by Borenstein et al. (2009) "if the missing studies are a random subset of all relevant studies, the failure to include these will certainly result in less information, wider confidence intervals, and less powerful tests, but will have *no systematic impact on the effect size*" (p. 277, italics added).

When research is systematically, not randomly, omitted from a synthesis, however, there is a publication bias because "readers and reviewers of that research are in danger of drawing the wrong conclusion about what that body of research shows" (Rothstein et al., 2005, p. 1). Thus, the extant meta-analytic literature reflects an enduring and ubiquitous concern regarding the universality and consequences of the file drawer phenomenon on the synthesis of research (Borenstein et al., 2009; Howard et al., 2009; McDaniel, Rothstein, & Whetzel, 2006; Rothstein et al., 2005; Sutton, 2009). The basic notion of the file drawer problem is the contention that research with statistically nonsignificant results is less likely to be published (e.g., see Borenstein et al., 2009; Cooper, 2010; Rothstein et al., 2005; Sutton, 2009; see also, Walster & Clearly, 1970 for an early discussion of this issue). Consider the summary of the file drawer problem provided by Cooper (2010): "research published in many journals is more likely to present statistically significant findings—that is, findings that reject the null hypothesis with a probability of $p < .05$ (or some other significance criterion)—than all research on the topic. This bias against null findings is present in the decisions made by both reviewers and primary researchers" (p. 62).

As noted by Rothstein et al. (2005), there has been publication bias through the censorship of studies for as long as research has been conducted and reported. Attention to that phenomenon, however, has increased in recent years largely with the widespread adoption of meta-analytic approaches to summarize research. Indeed, from its earliest foundations (e.g., Bakan, 1967; Cooper, 1979; McNemar, 1960; Orwin, 1983; Rosenthal, 1979; Smith, 1980; Sterling, 1959), the file drawer problem and related issues are easily among the most extensively chronicled critiques of the responsible interpretation of meta-analyses, or their lack thereof (Becker, 2005; Dickerson, 2005; Duval, 2005; Hedges & Vevea, 1996; Rothstein & Busing, 2005; Rothstein et al., 2005; Schwarzer, Antes, & Schumacher, 2003; Sterne, Becker, & Egger, 2005; Sutton, 2005, 2009).

In fact, the general criticism that meta-analyses do not include a random sample (i.e., an unbiased set of all available effect sizes) prompted Bonett (2008, 2009) to call into question the use of the most widely accepted and implemented random effects meta-analytic methods including the Hunter-Schmidt (2004) and Hedges and Vevea (1998) models. Such a suggestion is consequential in as much as a review of meta-analyses published in five OBHRM, I-O psychology, and management journals over the period 1982 through 2009 ascertained that more than 80% of reported effect sizes were computed using the Hunter-Schmidt procedures (Aguinis, Dalton et al., 2011).

A series of ingenious developments have been designed to ascertain whether a meta-analysis includes as many statistically nonsignificant results as one would expect from a given array of effect sizes (see Sutton, 2009 for an extensive summary; see also, Kromrey & Rendina-Gobioff, 2006). These include "fail-safe N" approaches (Becker, 2005; Cooper, 1979; Gleser & Olkin, 1996; Orwin, 1983; Rosenthal, 1979), nonparametric/rank correlation tests (Begg & Berlin, 1988; Begg & Mazumber, 1994), linear-regression tests (Egger, Davey-Smith, Schneider, & Minder, 1997; Stern & Egger, 2005), funnel plots (Light & Pillemer, 1984; Macaskill, Walter, & Irwig, 2001; Stern et al., 2005), and the trim and fill method (Duval, 2005; Duval & Tweedie, 2000a, 2000b; Schwarzer, Carpenter, & Rucker, 2010). In other words, there is an extensive methodological body of work developed with the specific objective of mitigating the supposedly important biasing effects of the file drawer problem.

Curiously, however, even with this distinguished body of work, there remains a conspicuous frustration that endures from the work of Rosenthal (1979) to present. According to the Google Scholar database, Rosenthal's (1979) discussion of this phenomenon has been cited 1,805 times as of October 14, 2011. As noted, we can determine that the elements of a given data array are potentially biased in that they do not include as many statistically nonsignificant results as would be expected. For example, we might know that a given array is expected to have five to eight more correlations that are not statistically significant. What we do not know, however, is whether 10%, 20%, 50%, or more of the entire body of reportable research relevant to any given synthesis is nonpublished (e.g., Rothstein et al., 2005; Sutton, 2005). If the general premise of the file drawer problem is correct, we simply do not know how many statistically nonsignificant results there are. Without that information, the validity of the entire body of published research—primary research, narrative reviews, and meta-analyses—may be compromised (Fiedler, 2011). Once again, there is some direct and derivative evidence of the existence of publication bias (e.g., Rothstein et al., 2005; Stern & Egger, 2005; Sutton, 2005, 2009), but there is no estimate of its magnitude. This sentiment is captured by Cooper and Hedges (2009): "As might be expected, publication bias is easier to detect than to correct, at least to correct in a way that inspires great confidence" (p. 566). Given the enduring attention to the file drawer problem and interest in the magnitude, if any, of its impact on meta-analysis and substantive theory and practice, we propose a novel approach, which we describe next.

File Drawer Problem Through an Alternative Lens

Consider the analytical outcome of a contemporary journal article in OBHRM, I-O psychology, and related fields. That estimate of a given statistical outcome is not likely to be presented as a bivariate correlation—the simple relationship of one variable (X) to another (Y). More likely, the analysis/analyses that address some hypothesis or hypotheses of interest will be multivariate in form, or the test of a model, or a factor analysis, or any of a host of statistical procedures that go beyond a single bivariate correlation analysis (Aguinis, Pierce, Bosco, & Muslin, 2009). In any case, where, then, does a synthesist find the data that will constitute the input data for a meta-analysis, particularly in the case of nonexperimental research? The answer is largely from a correlation matrix, which is a repository of bivariate relationships. Importantly, such matrices have several fascinating aspects for the meta-analyst. Consider, for example, a correlation matrix included in a hypothetical study as follows. This study focuses on a relationship of interest between some criterion variable Y and, for the sake of discussion, six predictor variables and two moderator variables. This analysis also has two control variables. All of these variables could be included in the test of a single model. What is imminently clear, however, is that these variables were chosen by the researcher for some analytical purpose, and these choices were presumably warranted on some theoretical bases. Given that, the resulting correlation matrix, comprised of 11 variables and, critically for the meta-analyst, bivariate correlations for each of their combinations, provides *r*s that are entirely appropriate as input for subsequent meta-analyses. Notably, for the study just described, none of these data are clutter. Each of the variables was specified—and in many cases hypothesized—to have some relationship with others. The number of effect sizes in this example is nontrivial. The number of unique effect sizes in a given correlation matrix is $k(k\text{-}1)/2$. In our 11-variable example, the number of unique effects sizes included in the correlation matrix would be 55.

Given the pervasiveness of electronic databases, the starting point for any meta-analytic review involves a search of relevant primary-level studies using certain keywords. However, conducting a meta-analysis based on an electronic search strategy exclusively is likely to lead to an incomplete set of relevant primary-level studies. Moreover, the recommendation that the search for effect sizes to be included in any given meta-analysis go beyond what seem to be directly relevant articles has been known and implemented for at least a quarter of a century.

McEvoy and Cascio (1985), for example, conducted a meta-analysis of antecedents of turnover and noted that "the search involved several literatures not normally associated with turnover research, such as supervisor training, behavior modeling, assessment centers, and weighted application blanks" (p. 343). Similarly, in a widely cited meta-analysis of antecedents and consequents of organizational commitment (OC), Mathieu and Zajac (1990) noted that "an article-by-article search of the Journal of Applied Psychology, the Academy of Management Journal, Administrative Science Quarterly, Human Relations, Organizational Behavior and Human Decision Processes, and Personnel Psychology was performed for the period January 1980 through September 1987. This final effort revealed several studies that were designed primarily to investigate other topics, yet included correlations with OC" (p. 172). In yet a third illustration of a more recently published meta-analysis, Zhao, Wayne, Glibkowski, and Bravo (2007) stated that "[b]ecause an electronic search with specific keywords may sometimes miss relevant studies. . . we also conducted a manual search of journals that regularly publish . . . [research that may be relevant to the meta-analytic investigation]" (p. 658).

Thus, all elements in published correlation matrices are likely candidates for inclusion in a meta-analysis even though the main research question and even research domain in the primary-level study may not be directly related to the main research question and research domain of the subsequent meta-analysis. This issue is highlighted by perusing appendices of meta-analyses that list the articles that were included in the review. For example, Williams, McDaniel, and Nguyen (2006) conducted a meta-analysis of the antecedents and consequences of pay satisfaction, but their meta-analysis included correlations extracted from published articles addressing a wide range of different topics and research domains. In fact, many of the articles that were used as sources for effect sizes addressed issues that are completely unrelated to the meta-analytic goal of investigating the relationship between pay satisfaction and its antecedents and consequences. These primary-level articles have titles such as "Lateness as a withdrawal behavior" (Adler & Golan, 1981), "Psychometric and substantive issues in scale construction and validation" (Drasgow & Miller, 1982), "A comparative study of mentoring among men and women in managerial, professional, and technical positions" (Dreher & Ash, 1990), and "Role of leadership in the employee withdrawal process: A constructive replication" (Ferris, 1985).

In short, relying on a literature search that focuses only on relevant keywords is known to result in a meta-analysis that does not include all relevant primary-level effect sizes because many studies include relevant variables without mentioning them in the title or abstract. The fact that so many meta-analyses make this point explicitly highlights the key point that *all* effect sizes included in published correlation matrices are relevant candidates for inclusion in subsequent meta-analyses even when some of the elements in a given correlation matrix are not central to the primary-level study in question.

Of potentially more interest for our present purposes, some of the *r*s included in a given correlation matrix may be statistically nonsignificant. That, however, is of absolutely no consequence. Indeed, statistically non-significant elements in correlation matrices may not be subject to whatever bias against statistically nonsignificant results has been broadly hypothesized. No forces conspired to prevent the publication of these correlation matrix data. We suspect that few journal submissions, which were otherwise in order, have been rejected on the basis of null correlations in their respective matrices. This is a potentially consequential observation for several reasons. Consider, for example, that many correlation matrices present us with hundreds of elements unfettered, we suspect, by any notion of selection bias. If one considers the number of available correlation matrices in the entire body of published research, the repository of bivariate correlations is vast. Moreover, all of these correlations are suitable for inclusion in meta-analyses.

As noted in an earlier section of our article, the file drawer problem is not just a missing-at-random data problem. Rather, the issue is whether research findings that appear in the published literature are unrepresentative of the population of completed studies, an unknown percentage of which were not published (e.g., Rothstein et al., 2005). Beyond that, at the foundation of the file drawer problem is the supposition that some group of unreported studies is not representative because these studies present far more often with statistically nonsignificant results (e.g., Borenstein et al., 2009).

It is on that basis that we suggest an alternative lens by which to examine the file drawer problem. Accordingly, we suggest that an assessment of the magnitude of the file drawer problem should be based on the equivalency, or lack thereof, of the correlation matrices of published and nonpublished research. Given that, if the pivotal issue is whether the correlations reported in nonpublished studies are fundamentally different in terms of statistical significance and magnitude compared to those reported in published studies, then an appropriate way in which to determine the accuracy of such allegations is to study the actual source of these data. And, that source is not a single bivariate effect size that as we have noted, rarely exists in that form in published and, we presume, nonpublished research. Rather, a more robust source for such an assessment will be the correlation matrices.

To provide an assessment of the extent of the file drawer problem, we derive estimates as follows: (a) frequency of statistically nonsignificant results reported in primary-level studies; (b) the extent to which published meta-analyses rely on such statistically nonsignificant results; (c) the extent to which the frequency of statistically nonsignificant elements in the correlation matrices in published primary-level studies is similar to the frequency of statistically nonsignificant elements in the correlation matrices of nonpublished primary-level studies, and (d) the magnitude of correlations reported in published research versus doctoral dissertations. Next, we describe five studies designed to obtain these estimates.

Study 1: Percentage of Statistically Nonsignificant Correlations Reported in Primary Studies

The goal of Study 1 was to obtain an estimate of the frequency and percentage of statistically nonsignificant correlations that are included in published correlation matrices from nonexperimental research. Data for this study consisted of a sample of correlation matrices reported in *JAP, PPsych,* and *AMJ*. The choice of these journals was based on three criteria. First, they are among the highest echelon of scholarly journals in OBHRM and I-O psychology in terms of reputation as well as impact factor (Cascio & Aguinis, 2008). Second, articles published in these three journals routinely include correlation matrices that, as noted, are critical for our analyses. Third, given the large number of submissions received by each of these journals as well as the highly rigorous and selective review process, these journals have some of the lowest acceptance rates in OBHRM, I-O psychology, and related fields. Accordingly, it seems reasonable to infer that the file drawer problem would be most likely to be observed in such research outlets.

Method

We randomly selected issues of *JAP, PPsych,* and *AMJ*, using Excel's random number generator from the following three periods: 1985–1989, 1995–1999, and 2005–2009. We chose the 1985–2009 period because we wanted to include articles published in the past quarter century, which begins 6 years after the publication of Rosenthal's (1979) article on the file drawer problem and ends when we began data collection (i.e., 2010). Once we chose our 25-year range, we selected 5-year periods evenly spaced to avoid potential selection bias. For each of the selected issues (e.g., *JAP*, 1997, issue 2), we examined the correlation matrices for all articles in that issue and extracted information regarding two variables: (1) size of the correlation matrix and (2) number and percentage of effect sizes that were not statistically significant. We used the .05 cutoff as the criterion for statistical significance, which required no judgment on our part. As reported in each article, a given correlation in the matrix of interest was either significant at the .05 level or it was not. Whether a given correlation was "substantive" or "of practical consequence" or equivalently described was not a factor.

Results and Discussion

We conducted the following analyses. First, we computed the percent of correlations that are statistically nonsignificant in each of the three journals. To do so, we computed the percent of statistically nonsignificant correlations in each matrix and then produced an average of these percentages across all matrices for each journal. This type of analysis at the matrix level considers the possibility that larger matrices are more likely to include more statistically significant correlations than smaller matrices. Second, we computed confidence intervals around the mean average percentages for each of the three journals. Third, we computed the correlation between the size of each matrix and the percent of nonsignificant effect sizes combining all matrices for each of the three journals. Studies reporting larger correlation matrices are likely to include more correlations that are not central to the substantive issues under examination by the original researchers so may be less likely to be rejected by journal editors who are not concerned about these nonsignificant effect sizes. Thus, this third analysis provides useful information because a positive correlation might provide evidence of the file drawer phenomenon. Fourth, we computed the average percentage of statistically nonsignificant correlations across all matrices across the three journals. Fifth, we computed the average sample size across all matrices for each journal together with 95% confidence intervals. As we describe in the General Discussion section, this information is important to understand whether the magnitude of correlations, and not only the percentage of statistically nonsignificant correlations, is similar across published and nonpublished sources.

For *JAP*, our analyses included 149 matrices with a total of 15,203 correlations. Across the 149 matrices, 46.51% of the correlations were statistically nonsignificant. The 95% confidence interval (CI) around the mean ranged from 42.71% to 50.31%. The correlation between matrix size and percentage of statistically nonsignificant correlations was $r(147) = .046$, $p = .57$. Finally, the correlation matrices from *JAP* upon which we based our analyses are based on a mean sample size of 286.93 with a 95% CI from 226.11 to 347.75. The median sample size was 198, which is comparable to the median sample size of 173 reported by Shen et al. (2011) for all articles published in *JAP* from 1995 to 2008. The similarity in sample size between our results and those reported by Shen et al. (2011) provides additional evidence regarding the representativeness of our data, which we collected by implementing a random procedure for the selection of journal issues.

For *PPsych*, our analyses included 79 matrices with a total of 10,217 effect sizes. Across the 79 matrices, 46.53% of the correlations were statistically nonsignificant and the 95% CI ranged from 41.14% to 51.92%. The correlation between the size of the correlation matrices and percentage of nonsignificant correlations was $r(77) = .037$, $p = .75$. Across all the *PPsych* correlation matrices, these results are based on a mean sample size of 301.24 with a CI of 234.57 to 367.91. The median sample size was 204.

For *AMJ*, our analyses were based on 175 matrices including 12,550 correlations. Across these matrices, the average of statistically nonsignificant correlations was 48.08% and the 95% CI ranged from 44.89% to 51.27%. The correlation between the size of the correlation matrices and the percent of nonsignificant effects resulted was $r(173) = .062$, $p = .41$. These estimates of the *AMJ* data are based on a mean sample size of 243.49 (95% CI of 203.42–283.56). The median sample size was 161.

Taken together, results based on a total of 37,970 effect sizes reported in 403 correlation matrices in *JAP*, *PPsych*, and *AMJ* indicate that the percentage of statistically nonsignificant correlations available to meta-analysts is quite similar across the three journals and also quite large—not remotely near zero. Regarding the similarity of results across journals, averages are 46.51% for *JAP*, 46.53% for *PPsych*, and 48.08% for *AMJ*. Moreover, evidence regarding the similarity of these mean percentages is that the 95% CIs overlap considerably. Given the similarity of results across journals, we combined the 403 matrices to compute an overall percentage of statistically nonsignificant correlations and found a mean of 46.81% (95% CI of 44.58%–49.05%).

We acknowledge that to demonstrate that so many statistically nonsignificant correlations are available for potential inclusion in subsequent meta-analyses does not necessarily mean that a similarly high percentage of nonsignificant correlations is actually used by published meta-analyses. The determination of whether the percentage of statistically nonsignificant correlations available from primary-level studies is similar to the percentage of statistically nonsignificant correlations used as input in published meta-analyses is the goal of Study 2.

Study 2: Percentage of Statistically Nonsignificant Correlations Used in Published Meta-Analyses

The goal of Study 2 was to generate an estimate of the frequency and percentage of statistically nonsignificant correlations from primary-level studies used as input in published meta-analyses. As previously noted, the issue of statistically nonsignificant results and its potential impact has largely been associated with meta-analyses. In Study 1, we found that 46.81% of published correlations available for future meta-analyses are statistically nonsignificant. In Study 2, we used correlations included as input in published meta-analyses to ascertain whether and to what extent published meta-analyses may have relied on a much smaller percentage of statistically nonsignificant correlations, which would signal the presence of the file drawer problem.

Method

Aguinis, Dalton et al. (2011) content analyzed 196 meta-analyses published in *AMJ, JAP, PPsych,* and other organizational science journals from January 1982 through August 2009. Of these 196 published meta-analyses, 51 included a list of the primary-level studies from which effect sizes were extracted and meta-analyzed (see Aguinis, Dalton et al., 2011, for a detailed description of their database). The data we used in Study 2 included 6,935 primary-level correlations that were used in those 51 published meta-analyses.

Results and Discussion

Results indicate that, across the 51 published meta-analyses, 44.31% (95% CI of 39.38%–49.24%) of the correlations are statistically nonsignificant. These estimates are based on a mean sample size of 320.80 with a 95% CI of 264.69–376.91. The median sample size was 326.5.

The result of 44.31% we obtained across the 51 published meta-analyses is similar to the percentage of 46.81% obtained across the 403 correlation matrices reported in published primary-level studies described in Study 1. Moreover, the 95% CIs around these two mean percentages overlap considerably. Accordingly, results provide evidence that there is a repository of statistically nonsignificant results in primary-level published studies that is similar to the percentage of statistically nonsignificant effect sizes that comprise the primary level input data for published meta-analyses.

Study 3: Percentage of Statistically Nonsignificant Correlations Reported in Nonpublished Studies

Even with the notable aspect of the similarity of the estimates obtained in Study 1 and Study 2, there remains yet a third estimate that may facilitate our understanding of the extent of the file drawer problem. Specifically, results from Studies 1 and 2 are about primary-level correlations from published sources. We do not know the nature of the effect sizes in nonpublished data, and it is that information that is crucial to estimate the magnitude of the file drawer problem. As previously noted, the basis of the file drawer problem is that statistically nonsignificant correlations are less likely to be published and thus will not be included in subsequent meta-analyses. Given our analyses thus far in Studies 1 and 2, an unanswered question is the extent to which correlation matrices in nonpublished primary-level studies are distinctly different from their published counterparts. Accordingly, the goal of Study 3 was to generate an estimate of the frequency and percentage of statistically non-significant correlations reported in nonpublished correlation matrices for nonexperimental research.

Method

To obtain correlation matrices from nonpublished manuscripts, we contacted the faculty members of 30 schools of business and 30 I-O psychology programs using the SIOP and *Business Week* directories (Society for Industrial and Organizational Psychology [www.SIOP.org/gtp]; *Business Week* [www.businessweek.com/business-schools]). We randomly selected one half of the listed faculty members at any given institution and personally contacted them. We used Excel's random number generator to implement our selection process (details about this procedure are available from the authors upon request).

The random selection process resulted in 361 faculty members. In an e-mail, we contacted these faculty members and requested "the correlation matrices for papers that you have decided over the years not to submit for publication or for whatever reasons have not been published." Those contacted were also informed that our interest was in the "file drawer problem in meta-analysis." Those contacted were also guaranteed that the collected data would be used without attribution to the "owner" or his or her institution. Moreover, those contacted were assured that we, having secured the requested data (e.g., the size of the correlation matrix and the nature of its effect sizes), would shred hard copies of the materials and delete the digital counterparts. Examples of universities in our sample include University of Florida (Management), Indiana University (Management and Entrepreneurship), Tulane University (I-O psychology), and Michigan State University (I-O psychology). In addition, the final selection of faculty included individuals at the assistant, associate, and full professor rank. We are not able to provide additional identifying information due to confidentiality concerns.

Results and Discussion

We received 167 correlation matrices (in some cases, more than one matrix was provided by a single respondent). For this study, as for Studies 1 and 2, we used the .05 cutoff for statistical significance as noted in each manuscript. Our data collection effort resulted in 13,943 effect sizes from these 167 correlation matrices. We found that 45.45% of the correlations were statistically nonsignificant, with a 95% CI of 42.29%–48.61%. The correlation between the size of the correlation matrices and percentage of nonsignificant effects was $r(165) = -.056, p = .47$. All of these estimates are based on a mean sample size of 290.11 with a 95% CI of 221.58%–358.64%. The median sample size was 197.

Results from Studies 1 through 3 indicate that the percentage of statistically nonsignificant correlations is similar across (a) those reported in primary-level studies published in *JAP, PPsych,* and *AMJ* (Study 1); (b) those used as input for meta-analyses published in *JAP, PPsych, AMJ,* and other organizational science journals (Study 2); and (c) those included in non-published primary-level studies (Study 3). Next, we describe an additional study with the goal to produce further evidence regarding the extent of the file drawer problem.

Study 4: Percentage of Statistically Nonsignificant Correlations Reported in Doctoral Dissertations

The goal of Study 4 was to obtain an estimate of the frequency and percentage of statistically nonsignificant correlations reported in nonexperimental doctoral dissertations. A comparison of correlations reported in dissertations to those reported in published studies is informative for two reasons. First, unlike published studies, dissertations are available regardless of whether they report statistically significant or nonsignificant correlations. Second, nonexperimental dissertations often contain a large number of variables without regard to statistical significance or hypotheses. In contrast, it is common practice that publications based on nonexperimental dissertations only report a subset of all the variables measured, typically those that are supportive of a subset of the dissertation hypotheses. Because authors of dissertations can choose to publish a subset of results, dissertations may contain a greater number of correlations compared to the number reported in publications based on the same research. Consequently, the percentage of nonsignificant correlations reported in dissertations may differ from that reported in published studies (Shadish, Doherty, & Montgomery, 1989; Smith, 1980).

Method

We obtained correlation matrices from doctoral dissertations by locating full-text dissertations via the ProQuest Dissertations and Theses A&I database from the same time periods examined in Study 1 (i.e., 1985–1989, 1995–1999, and 2005–2009). To obtain a sample with a similar research domain as those described in Study 1, we included dissertations with micro-area subject classifications such as "management and organizational behavior," "management and human resources," and "occupational psychology." Beyond these inclusion criteria, dissertations must have lacked explicit mention of an experimental research methodology and included at least one correlation matrix. Although most dissertations meeting these inclusion criteria were authored by individuals receiving doctoral degrees from management or psychology departments, departmental affiliation played no role in our inclusion criteria.

Our search resulted in locating 17 full-text dissertations for 1985–1989, 314 full-text dissertations for 1995–1999, and 956 full-text dissertations from 2005–2009. To expand our sample for the 1985–1989 time period, we relaxed our search criteria to include only the subject "management," excluding the Boolean AND operator, which resulted in an additional six for a total of 23 dissertations. We used random selection without replacement for the 1995–1999 and 2005–2009 time periods until 50 dissertations containing at least one correlation matrix were retrieved for each of these two time periods. In sum, our final sample consisted of 217 correlation matrices reported in 123 dissertations.

Results and Discussion

Results indicate that, across the 217 matrices including 20,860 correlations, 50.78% of the correlations are statistically nonsignificant. The 95% CI around the mean ranged from 47.24% to 54.32%. The correlation between matrix size and percentage of statistically nonsignificant correlations was $r(215) = .029$, $p = .67$. Finally, the 217 matrices were based on a mean sample size of 236.17 (median = 127.00) and the 95% CI around the mean ranged from 167.53 to 304.81.

As described earlier, doctoral dissertations are less likely to be affected by the file drawer problem compared to papers published in highly selective peer-reviewed journals. Moreover, doctoral dissertations are likely to include more variables and thus more correlations compared to published papers, including some that are not directly linked to central hypotheses. Nevertheless, our results show that the percentage of statistically nonsignificant correlations is quite similar to those obtained in Study 1, which was based on articles published in *JAP, PPsych,* and *AMJ*. Moreover, the 95% CI around the mean percentage in Study 4 overlaps with the CIs for each of the data sources we examined in Studies 1 through 3.

Study 5: Magnitude of Correlations Reported in Primary Studies Versus Doctoral Dissertations

Studies 1 through 4 focused on the frequency and percentage of statistically versus nonstatistically significant correlations. As we describe in more detail in the General Discussion section, results based on Studies 1 through 4 provide indirect evidence regarding lack of differences in the magnitude of correlation coefficients comparing published versus nonpublished effect sizes. The goal of Study 5 was to gather direct evidence regarding a possible difference in the average magnitude of correlations reported in published studies versus doctoral dissertations. Similar to the impetus for Study 4, dissertations are available regardless of whether they report statistically significant or nonsignificant correlations and, therefore, if the file drawer problem exists, the magnitude of correlations reported in primary studies should be larger than the magnitude of correlations reported in doctoral dissertations.

Method

We randomly sampled correlation matrices used in Study 1 (studies published in *JAP, PPsych,* and *AMJ*) and correlation matrices used in Study 4 (doctoral dissertations) until we reached at least 1,000 correlations for each of the two samples. In other words, we randomly selected one matrix out of 403 for Study 1 and extracted all correlations and the sample size upon which the correlations were computed. Then, we randomly selected a second matrix and also extracted the correlations and sample size and continued with the process until we had extracted at least 1,000 correlations. We used the same process to randomly select correlations matrices out of the total of 217 used in Study 4. Once the data collection was complete, we computed a within-matrix average and then a weighted (by sample size) average correlation for each of the two samples. In addition, we also computed confidence intervals around each of the weighted averages as well as a chi-square distributed Q_B statistic. The Q_B statistic provides information on whether there is a difference between the mean effect sizes across samples greater than would be expected from sampling error alone. In other words, Q_B tests the hypothesis that source of data (i.e., published studies vs. doctoral dissertations) is a moderator variable that has an effect on the magnitude of the obtained correlations (Aguinis, Gottfredson, & Wright, 2011; Aguinis, Sturman, & Pierce, 2008).

Results and Discussion

The sample based on data from published studies included 1,002 correlations from nine matrices (i.e., about 111 unique correlations per matrix) and the sample based on doctoral dissertations included 1,224 correlations from 11 matrices (i.e., also about 111 unique correlations per matrix). Given that the number of unique elements in a correlation matrix is determined by $k(k\text{-}1)/2$, each matrix had, on average, a dimension of about 14 by 14. The average dimension of matrices in Study 1 and Study 4 was between 14×14 and 15×15. Thus, the similarity in the average dimension of the matrices provides evidence regarding the representativeness of the samples of matrices and correlations used in Study 5.

We first obtained an average effect size within each correlation matrix. Then, we computed a mean weighted (by sample size) correlation across each of the two sets of matrices. For the published studies, the mean weighted (by sample size) correlation was .2270 and the 95% confidence interval ranged from .1779 to .2762. For the doctoral dissertations, the mean weighted (by sample size) correlation was .2279 and the 95% confidence ranged from .1819 to .2740. In addition, Q_B $(df = 1) = .0005$, $p = .98$, indicating that the magnitude of the correlations does not differ across samples. We also conducted the same analysis but computed the mean weighted (by sample size) effect sizes based on the 1,002 correlations from Study 1 and the 1,224 correlations from Study 4, and substantive results remained the same (i.e., means of .225 and .226, respectively, and Q_B $[df = 1] = .048$, $p = .83$).

The mean correlations are virtually identical across the two samples, there is overlap in the confidence intervals, and the Q_B test indicates homogeneity of correlations across samples. Although we were surprised by the near identical values for the mean correlations, we anticipated their approximate magnitude because Aguinis, Dalton et al. (2011) found a median correlation of .23 based on 5,581 meta-analytically derived correlations. Aguinis, Dalton et al. (2011) extracted the 5,581 meta-analytically derived correlations from 196 published meta-analyses, most of which used both published and nonpublished correlations as inputs. So, the data reported by Aguinis, Dalton et al. (2011) as well as the data we used in Study 1 and Study 4 presumably were drawn from the same super population of primary-level data available in the entire field (published and unpublished). This explains the similarity regarding mean correlations across all three sources of data (i.e., Study 1 using published primary-level studies; Study 4 using doctoral dissertations; and Aguinis, Dalton et al. using published meta-analyses).

In sum, results indicate that there is no substantive difference in the magnitude of correlations reported in primary studies compared to those reported in doctoral dissertations. Accordingly, Study 5 provides additional evidence negating the putative effect of the file drawer problem.

General Discussion

The file drawer problem is one of the most enduring threats to the validity of meta-analytic conclusions. The assumption is that null results are less likely to be published compared to statistically significant results and, hence, less likely to be included in meta-analytic reviews. Supposedly, this results in an upwardly biased sample of primary-level effect-size estimates and upwardly biased meta-analytically derived effect sizes. If the general premise of the file drawer problem is correct, then it is a cogent indictment for meta-analysis. More importantly, the meta-analytic summaries of entire bodies of research of the scholarly community are adulterated. Given that, we cannot accurately, irrespective of our well-intended efforts, interpret our meta-analytic findings and where they fit in the fabric of related work. Results based on the five studies reported herein are remarkably consistent. Figure 1 provides a visual summary of the point estimates as well as CIs for the percentage of statistically nonsignificant correlations obtained across each of the different data sources in Studies 1 through 4.

As shown in Figure 1, between 40% and 50% of primary level effect sizes in published and nonpublished sources that are potentially included and actually included in meta-analytic reviews are not statistically significant. We reached this conclusion by examining (a) 37,970 primary-level

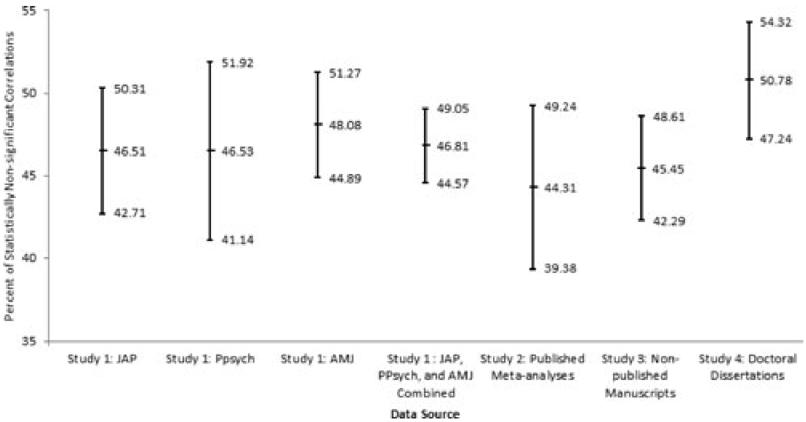

Figure 1: **Overlap of 95% Confidence Intervals Around Mean Percentages of Statistically Nonsignificant Correlations Across Studies 1–4.**

Note. *JAP*: Primary-level studies published in *Journal of Applied Psychology; PPsych*: Primary-level studies published in *Personnel Psychology; AMJ*: Primary-level studies published in *Academy of Management Journal*.

effect sizes reported in 403 correlation matrices published in *JAP, PPsych*, and *AMJ*; (b) 6,935 effect sizes used as input in 51 meta-analyses published in *JAP, PPsych, AMJ*, and other organizational science journals; (c) 13,943 correlations included in 167 matrices reported in non-published manuscripts authored by I-O psychology and management scholars; and (d) 20,860 correlations included in 217 matrices reported in doctoral dissertations.

The importance of the file drawer problem for theory and practice relies on the concern that meta-analytically derived effect sizes may be inflated. We acknowledge that our analyses for Studies 1–4 are based on the percentage of statistically non-significant correlations and these studies did not assess differences in magnitude between published and non-published correlations directly. However, the probability of finding a correlation that is statistically significant is determined by the following three factors: (1) pre-specified Type I error rate (α), (2) sample size, and (3) effect size (Cohen, 1988). For Studies 1–4, we held α constant at .05. In other words, a correlation was considered statistically significant if it reached the .05 threshold as reported in each study. Second, sample sizes across each of the four studies are similar. As reported earlier, for each of the four studies, we collected information on the sample sizes used to compute correlations. As a summary of these results, Figure 2 includes a graphic representation of each of the 95% confidence intervals around mean sample sizes across our data sets. As shown in this figure,

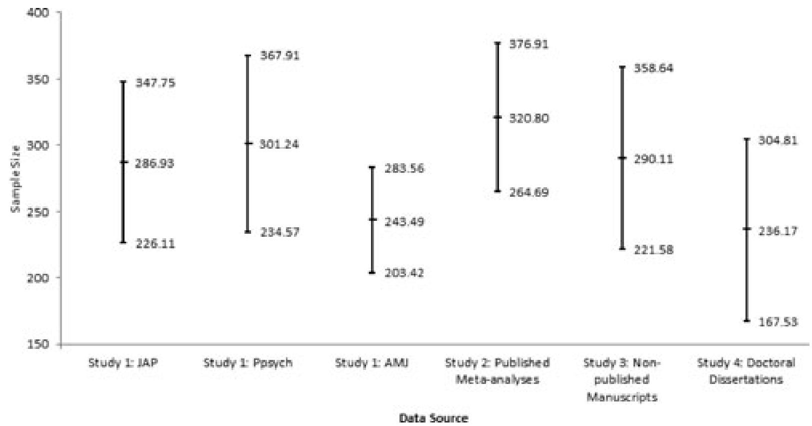

Figure 2: **Overlap of 95% Confidence Intervals Around Mean-Sample Sizes Across Studies 1–4.**

Note. *JAP*: Primary-level studies published in *Journal of Applied Psychology; PPsych*: Primary-level studies published in *Personnel Psychology; AMJ*: Primary-level studies published in *Academy of Management Journal*.

all of the confidence intervals overlap, indicating the sample size was similar across data sources. Finally, our results, which are summarized graphically in Figure 1, also show that there is overlap across all confidence intervals around mean percentages of statistically nonsignificant correlations across all data sets. Thus, given that there are no differences regarding the prespecified Type I error rates, sample size, and percentage of nonsignificant correlations across data sets, we can conclude that there are no differences in the magnitude of correlation coefficients comparing published versus nonpublished effect sizes. This indirect evidence was corroborated by direct evidence based on Study 5 regarding the similarity in the magnitude of correlations reported in primary studies compared to correlations reported in doctoral dissertations. Taken together, these results have implications for substantive theory and practice as well as methodological practices, which we describe next.

Implications for Theory and Practice

Our results suggest that, contrary to the established and long-held belief, the file drawer problem does not pose a serious threat to the validity of meta-analytically derived conclusions. Based on the belief that the file drawer problem is of great concern, authors of meta-analyses routinely take steps to address it when discussing implications for substantive conclusions. For example, many authors issue warnings that, due to assumed effects of the file drawer problem, substantive conclusions should be interpreted cautiously—implying that the obtained meta-analytic effect sizes may be larger.

To further understand the magnitude of the file drawer problem, consider the following illustrations from articles published in *JAP* and *PPsych*. Rhoades and Eisenberger (2002) noted that "the majority of antecedents and all of the consequences passed the 5k + 10 guideline (Hedges & Olkin, 1985; Rosenthal, 1979) wherein the fail-safe N should be larger than five times the number of studies included in the meta-analysis plus 10. Education, gender, and Conscientiousness were the only variables that failed to satisfy the guideline; *the statistically reliable relationships reported for these three variables should therefore be considered cautiously*" (p. 703, italics added). Also highlighting the magnitude of the file drawer problem, Tett and Meyer (1993) wrote that "based on the assumption that unpublished findings are weaker than (but as valid as) published findings (due to possible bias favoring acceptance of studies reporting stronger findings), and the fact that unpublished findings were largely excluded from the present aggregations, *current estimates of rho may be inflated*. Difficulty in obtaining the required information (i.e., K, mean N, mean r, and the relative validity of unpublished studies) *precludes firm judgments of the degree of overestimation*" (p. 267, italics added). Similarly, Allen, Eby, Poteet, Lentz, and Lima (2004) concluded that "in some cases, initially small effect sizes coupled with a small number of primary studies resulted in a small number of additional studies averaging null results that would be needed to reduce the effect size to .01 (e.g., mentoring and intentions to stay). *The reliability of the results pertaining to those relationships should be viewed cautiously*" (p. 132, italics added).

Our results suggest that these and many other similar caveats and cautionary statements about the file drawer problem do not seem to be warranted. This is indeed good news for substantive theory and practice because our results indicate that the concern that the file drawer problem leads to an overestimation of meta-analytically derived effect sizes is not warranted. In short, we found no evidence to support the long-lamented belief that the file drawer problem produces an upward bias in meta-analytically derived effect sizes (i.e., *r*s).

Implications for Methodological Practices

A common methodological practice is to discuss the implications of the file drawer problem in terms of results being robust and trustworthy. Specifically, if a meta-analysis "passes" a file drawer problem test, then substantive conclusions are interpreted more forcefully. Consider the following illustrations. Willness, Steel, and Lee (2007) noted that "in order to address potential 'file drawer' problems, Failsafe-N values were calculated for each of the variables, which estimates the number of unpublished studies with an average effect of zero that would be required to reduce a given meta-analytic coefficient to ± .10 (i.e., a small correlation with lower practical significance, as per Cohen, 1969). These results appear in Table 4, demonstrating that these findings are unlikely to be significantly affected by publication bias" (p. 143). Similarly, Steel and Ovalle (1984) concluded that "a total of 73,415 unpublished studies containing null results would be required to invalidate the present study's conclusion that behavioral intentions and employee turnover are significantly related" (p. 681). Finally, an article published in the July 2011 issue of *JAP* serves as an illustration of the relevance and currency of the file drawer problem. Specifically, Hartnell, Ou, and Kinicki (2011) included a table listing failsafe *k*s for each of the 25 meta-analytically derived correlations. The reason they did this is that they wanted to evaluate the "robustness of their findings given that . . . meta-analytic results should consider the 'file drawer' problem . . . because effect sizes may be biased" (p. 688).

Our results indicate that the methodological practice of estimating the extent to which results are not vulnerable to the file drawer problem may be eliminated. Stated differently, if, as our data suggest, the file drawer problem is in fact not producing a bias, then there is no need to "fix" an upward bias that does not exist.

We emphasize that we have based our analyses on correlation matrices in OBHRM, I-O psychology, and related fields. However, these conclusions are of no consequence for fields of research that rarely include a correlation matrix in the recitation of research results or for journals that do not require or strongly suggest them. Consider, for example, that the empirical work of researchers in fields such as accounting, finance, sociology, and economics rarely includes correlation matrices for the relevant variables. Perhaps this helps us understand why a computer-aided search of meta-analyses from 1980–2010 in the *Journal of Finance*, *Journal of Financial Economics*, *Accounting Review*, *Journal of Accounting Research*, *American Sociological Review*, *American Journal of Sociology*, *Rand Journal of Economics*, and the *American Economic Review* resulted in only two citations (i.e., Card & Krueger, 1995; Trotman & Wood, 1991).

An additional methodological implication of our results is that they can help bridge the much lamented micro–macro divide (Aguinis, Boyd, Pierce, & Short, 2011; Hitt, Beamish, Jackson, & Mathieu, 2007) by the suggestion to extend the methodological practice in OBHRM and I-O psychology journals to routinely include full correlation matrices in other fields. We firmly believe that extant matrices constitute a crucial repository for future research synthesis. An even larger repository, particularly in areas that have not required them historically, or even advocated, promises more inclusion of relevant data for such future syntheses. The intradisciplinary publication bias (e.g., Borenstein et al., 2009; Cooper, 2010), whereby meta-analysts may be more likely to include work from their own disciplines could also, at least in part, be addressed with the addition of correlation matrices originated in other fields of inquiry. Consider, for example, that research in accounting, finance, economics, and sociology could complement the body of research in strategic management studies, which, in turn, also relies on OBHRM and I-O psychology research at the individual and group level of analysis (Dalton & Dalton, 2005).

Potential Limitations and Suggestions for Future Research

Our results show that correlation matrices for published nonexperimental research are essentially equivalent to correlation matrices for nonpublished, nonexperimental research with regard to the percentage of nonsignificant effect sizes as well as the average size of effects. We have not, however, established this phenomenon at the focal level. Our data do not provide an insight into whether such comparisons would maintain for studies—published and nonpublished—particularly focused on, for example, the Big Five personality traits or employee withdrawal behaviors (e.g., absenteeism, transfers, and turnover). Our research could not examine this focal issue for two reasons. As noted in an earlier section, we requested, with the promise of anonymity, the correlation matrices of unpublished work from faculty members in management and I-O psychology. Thus, we could not determine which of the variables in such matrices were dependent, independent, control, moderator, or otherwise. In addition, even if we could have determined that information, we would not have had sufficient statistical power to analyze such data. For the data collection on which we relied, it would have been unlikely to have 15 cases, for sake of discussion, with *r*s for the relationship between a given X-Y for published data and 15 cases for that same X-Y for unpublished data. This power issue will be manifest for research of this type and can only be overcome by extensive data collection.

As an additional limitation to consider, our review focused on *r* as the effect-size index. Our choice was guided by the fact that the vast majority of meta-analyses published in OBHRM and I-O psychology are based on *r*s (Aguinis, Dalton et al., 2011). Thus, we are not necessarily able to extrapolate results from our analyses based on *r*s to meta-analyses based on other effect-sizes indices such as *d*s, which are more common in fields that favor the use of experimental design such as the medical sciences. Accordingly, future research can examine the degree to which the file drawer problem may affect meta-analytic conclusions derived from analyses based on effect-size indices other than *r*.

Conclusion

We conducted five studies to determine the existence and scale of the file drawer problem. Critically, there was essentially no difference in the percentage of statistically nonsignificant correlations reported in matrices in published compared to nonpublished studies. In addition, we gathered indirect and direct evidence in support of the similarity in the magnitude of published and nonpublished correlations. Much like the general population, researchers are not immune to received doctrines and things we just know to be true (Lance, 2011). These issues are "taught in undergraduate and graduate classes, enforced by gatekeepers (e.g., grant panels, reviewers, editors, dissertation committee members), discussed among colleagues, and otherwise passed along among pliers of the trade far and wide and from generation to generation" (Lance, 2011, p. 281). Our five studies provide consistent empirical evidence that the file drawer problem does not produce an inflation bias and does not pose a serious threat to the validity of meta-analytically derived conclusions as is currently believed.

REFERENCES

Adler S, Golan J. (1981). Lateness as a withdrawal behavior. *Journal of Applied Psychology*, *66*, 544–554.

Aguinis H. (2001). Estimation of sampling variance of correlations in meta-analysis. PERSONNEL PSYCHOLOGY, *54*, 569–590.

Aguinis H, Boyd BK, Pierce CA, Short JC. (2011). Walking new avenues in management research methods and theories: Bridging micro and macro domains. *Journal of Management*, *37*, 395–403.

Aguinis H, Dalton DR, Bosco FA, Pierce CA, Dalton CM. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, *37*, 5–38.

Aguinis H, Gottfredson RK, Wright TA. (2011). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior*, *32*, 1033–1043.

Aguinis H, Pierce CA, Bosco FA, Muslin IS. (2009). First decade of *Organizational Research Methods*: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, *12*, 69–112.

Aguinis H, Pierce CA, Bosco FA, Dalton DR, Dalton CM. (2011). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods*, *14*, 306–331.

Aguinis H, Sturman MC, Pierce CA. (2008). Comparison of three meta-analytic procedures for estimating moderating effects of categorical variables. *Organizational Research Methods*, *11*, 9–34.

Allen TD, Eby LT, Poteet ML, Lentz E, Lima L. (2004). Career benefits associated with mentoring for protégés: A meta-analysis. *Journal of Applied Psychology*, *89*, 127–136.

Bakan D. (1967). *Toward a reconstruction of psychological investigation*. San Francisco, CA: Jossey-Bass.

Becker BJ. (2005). Failsafe N or file-drawer number. In Rothstein HR, Sutton AJ, Borenstein M (Eds.), *Publication bias in meta-analysis* (pp. 111–125). West Sussex, England: John Wiley & Sons.

Begg CB, Berlin JA. (1988). Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society*, *151*, 419–463.

Begg CB, Mazumber M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088–1101.

Bonnett DG. (2008). Meta-analytic interval estimation for bivariate correlations. *Psychological Methods*, *13*, 173–189.

Bonnett DG. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods*, *14*, 225–238.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. (2009). *Introduction to meta-analysis*. West Sussex, England: John Wiley & Sons.

Burke MJ. (1984). Validity generalization: A review and critique of the correlational model. PERSONNEL PSYCHOLOGY, *37*, 93–115.

Burke MJ, Landis R. (2003). Methodological and conceptual issues in applications of meta-analysis. In Murphy K (Ed.), *Validity generalization: A critical review* (pp. 287–310). Hillsdale, NJ: Erlbaum.

Card D, Krueger AB. (1995). Time-series minimum-wage studies: A meta-analysis. *The American Economic Review*, *85*, 238–243.

Cascio WF, Aguinis H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, *93*, 1062–1081.

Cohen J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Cooper HM. (2010). *Research synthesis and meta-analysis* (4th ed.). Los Angeles, CA: Sage.

Cooper H, Hedges LV. (2009). Potentials and limitations. In Cooper H, Hedges LV, Valentine JC (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 561–572). New York, NY: Russell Sage Foundation.

Cooper H, Hedges LV, Valentine JC. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

Curran PJ. (2009). The seemingly quixotic pursuit of a cumulative psychological science. *Psychological Methods*, *14*, 77–80.

Dalton DR, Dalton CM. (2005). Strategic management studies are a special case for meta-analyses. In Ketchen DJ, Bergh DD (Eds.), *Research methodology in strategy and management* (Vol. 2, pp. 31–63). Boston, MA.: Elsevier.

Dickerson K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In Rothstein HR, Sutton AJ, Borenstein M (Eds.), *Publication bias in meta-analysis* (pp. 11–33). West Sussex, England: John Wiley & Sons.

Drasgow F, Miller HE. (1982). Psychometric and substantive issues in scale construction and validation. *Journal of Applied Psychology*, *67*, 268–279.

Dreher GF, Ash RA. (1990). A comparative study of mentoring among men and women in managerial, professional, and technical positions. *Journal of Applied Psychology*, *75*, 539–546.

Duval S. (2005). The trim and fill method. In Rothstein HR, Sutton AJ, Borenstein M (Eds.), *Publication bias in meta-analysis* (pp. 127–144). West Sussex, England: John Wiley & Sons.

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

Duval S, Tweedie R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

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

Ferris GR. (1985). Role of leadership in the employee withdrawal process: A constructive replication. *Journal of Applied Psychology*, *70*, 777–781.

Fiedler K. (2011). Voodoo correlations are everywhere—Not only in neuroscience. *Perspectives on Psychological Science*, *6*, 163–171.

Geyskens I, Krishnan R, Steenkamp J-BEM, Cunha PV. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management*, *35*, 393–419.

Gilboa S, Shirom A, Fried Y, Cooper C. (2008). A meta-analysis of work demand stressors and job performance: Examining main and moderating effects. PERSONNEL PSYCHOLOGY, *61*, 227–271.

Glass GV. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, *5*, 351–379.

Gleser LJ, Olkin I. (1996). Models for estimating the number of unpublished studies. *Statistics in Medicine*, *15*, 2493–2507.

Greenwald AG. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1–20.

Hartnell CA, Ou YA, Kinicki A. (2011). Organizational culture and organizational effectiveness: A meta-analytic investigation of the competing values framework's theoretical suppositions. *Journal of Applied Psychology*, *96*, 677–694.

Hedges LV, Olkin I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hedges LV, Vevea JL. (1996). Estimating effect size under publication bias in meta-analysis. *Journal of Educational and Behavioral Statistics*, *21*, 299–332.

Hedges LV, Vevea JL. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486–504.

Hitt MA, Beamish PW, Jackson SE, Mathieu JE. (2007). Building theoretical and empirical bridges across levels: Multilevel research in management. *Academy of Management Journal*, *50*, 1385–1399.

Howard GS, Lau MY, Maxwell SE, Venter A, Lundy R, Sweeny RM. (2009). Do research literatures give correct answers? *Review of General Psychology*, *13*, 1216–121.

Hunter JE, Schmidt FL. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). New York, NY: Academic Press.

Hunter JE, Schmidt FL, Jackson GB. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Kromrey JD, 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.

Lance CE. (2011). More statistical and methodological myths and urban legends. *Organizational Research Methods*, *14*, 279–286.

Le H, Oh I-S, Shaffer JA, Schmidt FL. (2007). Implications of methodological advances for the practice of personnel selection: How practitioners benefit from recent developments in meta-analysis. *Academy of Management Perspectives*, *21*, 6–15.

Light RJ, Pillemer DB. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

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

Mathieu JE, Zajac DM. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin*, *108*, 171–194.

McDaniel MA, Rothstein HR, Whetzel DL. (2006). Publication bias: A case study of four test vendors. PERSONNEL PSYCHOLOGY, *59*, 927–953.

McEvoy GM, Cascio WF. (1985). Strategies for reducing employee turnover: A meta-analysis. *Journal of Applied Psychology*, *70*, 342–353.

McNemar Q. (1960). At random: Sense and nonsense. *American Psychologist*, *15*, 295–300.

Mone MA, Mueller GC, Mauland W. (1996). The perceptions and usage of statistical power in applied psychology and management research. PERSONNEL PSYCHOLOGY, *49*, 103–120.

Mount MK, Oh I-S, Burns M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. PERSONNEL PSYCHOLOGY, *61*, 113–139.

Orwin RG. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157–159.

Rhoades L, Eisenberger R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology*, *87*, 698–714.

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

Rothstein HR, Busing A. (2005). Annotated bibliography of publication bias. In Rothstein HR, Sutton AJ, Borenstein M (Eds.), *Publication bias in meta-analysis* (pp. 331–346).West Sussex, England: John Wiley & Sons.

Rothstein HR, Sutton AJ, Borenstein M. (2005). *Publication bias in meta-analysis*. West Sussex, England: John Wiley & Sons.

Rousseau DM, Manning J, Denyer D. (2008). Evidence in management and organizational science: Assembling the full weight of scientific knowledge through synthesis. *The Academy of Management Annals*, *2*, 475–515.

Schmidt FL, Hunter JE. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*, 529–540.

Schmidt FL, Oh I-S, Hayes T. (2009). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, *62*, 97–128.

Schmidt FL, Shaffer JA, Oh I-S. (2008). Increased accuracy of range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. PERSONNEL PSYCHOLOGY, *61*, 827–868.

Schwarzer G, Antes G, Schumacher M. (2003). Statistical tests for the detection of bias in meta-analysis. In Schulze R, Holling H, Bohning D (Eds.), *Meta-analysis: New developments and applications in the biomedical and social sciences* (pp. 71–78). Seattle, WA: Hogrefe & Huber.

Schwarzer G, Carpenter J, Rucker G. (2010). Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. *Journal of Clinical Epidemiology*, *63*, 282–288.

Shadish WR, Doherty M, Montgomery LM. (1989). How many studies are in the file drawer? An estimate from the family/marital psychotherapy literature. *Clinical Psychology Review*, *9*, 589–603.

Shen W, Kiger TB, Davies SE, Rasch RL, Simon KM, Ones DS. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology*, *96*, 1055–1064.

Smith ML. (1980). Publication bias and meta-analysis. *Evaluation in Education*, *4*, 222–24.

Smith ML, Glass GV. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752–760.

Steel RP, Ovalle NK. (1984). A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover. *Journal of Applied Psychology*, *69*, 673–686.

Sterling TD. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*, 30–34.

Sterne JAC, Becker BJ, Egger M. (2005). The funnel plot. In Rothstein HR, Sutton AJ, Borenstein M (Eds.), *Publication bias in meta-analysis* (pp. 75–98). West Sussex, England: John Wiley & Sons.

Sterne JAC, Egger M. (2005). Regression methods to detect publication and other bias in meta-analysis. In Rothstein HR, Sutton AJ, Borenstein M (Eds.), *Publication bias in meta-analysis* (pp. 99–110). West Sussex, England: John Wiley & Sons.

Sutton AJ. (2005). Evidence concerning the consequences of publication and related biases. In Rothstein HR, Sutton AJ, Borenstein M (Eds.), *Publication bias in meta-analysis* (pp. 175–192). West Sussex, England: John Wiley & Sons.

Sutton AJ. (2009). Publication bias. In Cooper H, Hedges LV, Valentine JC (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435–452). New York, NY: Russell Sage Foundation.

Tett RP, Meyer JP. (1993). Job satisfaction, commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. PERSONNEL PSYCHOLOGY, *46*, 259–293.

Trotman KT, Wood R. (1991). A meta-analysis of studies on internal control judgments. *Journal of Accounting Research*, *29*, 180–193.

Viswesvaran C, Barrick MR, Ones DS. (1993). How definitive are conclusions based on survey data: Estimating robustness to nonresponse. PERSONNEL PSYCHOLOGY, *46*, 551–567.

Walster GW, Clearly TA. (1970). A proposal for a new editorial policy in the social sciences. *American Statistician*, *24*, 5.

Williams ML, McDaniel MA, Nguyen NT. (2006). A meta-analysis of the antecedents and consequences of pay level satisfaction. *Journal of Applied Psychology*, *91*, 392–413.

Willness CR, Steel P, Lee K. (2007). A meta-analysis of the antecedents and consequences of workplace sexual harassment. PERSONNEL PSYCHOLOGY, *60*, 127–162.

Zhao H, Wayne SJ, Glibkowski BC, Bravo J. (2007). The impact of psychological contract breach on work-related outcomes: A meta-analysis. PERSONNEL PSYCHOLOGY, *60*, 647–680.