

Publication Bias: A call for improved meta-analytic practice in the organizational sciences

George C. Banks, Sven Kepes and Michael A. McDaniel

Virginia Commonwealth University, 301 West Main Street, PO Box 844000, Richmond, VA 23284, USA.
banksgec@vcu.edu

Previous research has introduced the threat of publication bias to meta-analytic reviews in management and industrial/organizational (I/O) psychology research. However, a comprehensive review of top-tier journals demonstrates that more than two thirds of meta-analytic studies in management and I/O psychology ignore the issue. Of the studies that do empirically evaluate publication bias, almost all use methods that are based on problematic assumptions (e.g., the Failsafe N and subgroup comparisons by source of data). The current paper reviews the issue of publication bias and introduces to management and I/O psychology new methodological techniques to assess this bias. To illustrate the methods, multiple publication bias methods are demonstrated in a meta-analytic review of conditional reasoning tests for aggression. We offer specific recommendations that address both design and analysis issues to mitigate the existence and influence of publication bias.

1. Introduction

Publication bias exists to the extent that available research results are unrepresentative of all research results. The typical consequence of publication bias is an overestimation of effect sizes (Dickersin, 2005; McDaniel, Rothstein, & Whetzel, 2006). It has been proposed that publication bias is one of the greatest threats to the validity of meta-analytic reviews (Banks & McDaniel, 2011; Rothstein, Sutton, & Borenstein, 2005b), and that such reviews are one of our most important tools for advancing science and evidence-based management (Briner & Rousseau, 2011). Therefore, confidence in the validity and robustness of our meta-analytic results is contingent upon the extent to which publication bias influences our research.

Journal articles, books, and tutorials provide guidance on how to address the issue of publication bias in meta-analytic reviews (e.g., Berlin & Ghersi, 2005; Borenstein, Hedges, Higgins, & Rothstein, 2009; Hunter & Schmidt, 2004; McDaniel et al., 2006; Rothstein, Sutton, & Borenstein, 2005a; Song et al., 2010). Despite this, the majority of meta-analytic reviews published in the top-tier management and industrial/organizational (I/O) psychology

journals continue to ignore this type of bias. In this respect, our literature lags behind related psychological, educational, and medical research.

To demonstrate the degree to which our literature is methodologically behind other fields of research, we examined meta-analytic reviews published in top management and I/O psychology journals. This review complements previous reviews of publication bias (e.g., Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011a; Geykens, Krishnan, Steenkamp, & Cunha, 2009) by providing data from other fields of research as a comparison point. Additionally, our review complements and extends the work that was conducted by Ferguson and Brannick (2012), who obtained similar results in regard to general psychology. However, our review takes an additional step by illustrating that management and I/O psychology lags behind several fields in the evaluation and assessment of publication bias.

We reviewed the literature from 2005 to 2010 (we choose 2005 as our starting point as it coincides with the release of a key publication, *Publication bias in meta-analysis: Prevention, assessment, and adjustments* by Rothstein, Sutton, & Borenstein [2005a]). This period was thought to provide a representative sample of meta-

analytic reviews with a sufficient number of studies to have confidence in the identification of the different practices by the various fields.¹ We limited our review to top-tier journals that publish meta-analytic reviews with impact factors (IF) greater than 3.0 as reported in the 2009 Journal Citation Reports (2010). These journals were the *Academy of Management Journal*, the *Journal of Applied Psychology*, the *Journal of Management*, and *Personnel Psychology* (other top-tier journals such as the *Strategic Management Journal*, *Administrative Science Quarterly*, and *Organization Science* published less than five meta-analytic reviews and were excluded from our review).² For comparison purposes, we also reviewed top-tier journals (based on IF) in psychology, education, and medicine that have published at least five meta-analytic reviews between 2005 and 2010. These journals were *Psychological Bulletin* (IF = 12.85; although *Psychological Bulletin* publishes related research, it has been excluded from previous reviews of the management and I/O psychology literature; e.g., Podsakoff, MacKenzie, Podsakoff, & Bachrach, 2008), *Review of Educational Research* (IF = 3.33), and *Lancet* (IF = 30.76). These comparisons demonstrate how leading journals in other scientific disciplines evaluate publication bias. The results of this review are presented in Table 1.

Of the 105 meta-analytic reviews identified in our search of the top management and I/O psychology journals, only 33 (31%) evaluated publication bias empirically. Of those, the majority used either some form of the Failsafe N technique (Orwin, 1983; Rosenthal, 1979) or a moderator comparison by source (e.g., subgroup comparison of published vs. unpublished samples). Yet, it has been established that the former method, the Failsafe N technique, is inadequate to assess publication bias (Aguinis, Pierce, Bosco, Dalton, & Dalton, 2011b; Becker, 2005; Evans, 1996; McDaniel et al., 2006). Regarding the latter method, the evaluation of results by source type is also a limited technique when evaluating publication bias as it is restricted to comparing *identified* published samples to *identified* unpublished samples. Although this information can be interesting (i.e., to what extent do available journal articles differ from available conference papers, dissertations, etc.), it does not empirically evaluate the possibility that samples were not identified. Therefore, if one were to use a moderator comparison to infer whether or not publication bias is present, one makes the implicit assumption that 100% of all relevant published and unpublished samples have been identified. This assumption is rarely met and almost never tested (Hopewell, Clarke, & Mallett, 2005). Second, a moderator comparison by source type does not empirically evaluate the potential influence and magnitude of publication bias when taking into consideration the potential for missing studies.

In sum, the Failsafe N technique and subgroup comparisons by source are inadequate for assessing the

Table 1. Publication bias analyses from 2005 to 2010^a

Journal ^b	# of MAs	Emp. assessment of pb (%)	Source comparison (%)	Failsafe N (%)	Funnel plot ^c (%)	T&F (%)	CMA (%)	Egger's test (%)	B&M's test (%)	Other (%)	Multiple methods (%)
AMJ	8	6 (75)	3 (38)	3 (38)	0 (0)	1 (13)	0 (0)	0 (0)	0 (0)	0 (0)	1 (13)
JAP	62	19 (31)	11 (18)	8 (13)	1 (2)	1 (2)	0 (0)	0 (0)	0 (0)	0 (0)	2 (3)
JOM	9	2 (22)	1 (11)	1 (11)	0 (0)	1 (11)	0 (0)	0 (0)	0 (0)	0 (0)	1 (11)
PPsych	26	6 (23)	4 (15)	2 (8)	0 (0)	1 (4)	0 (0)	0 (0)	0 (0)	0 (0)	1 (3)
Total ^d	105	33 (31)	19 (18)	14 (13)	1 (1)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	5 (5)
Total: 2009–2010	35	11 (31)	6 (17)	4 (11)	0 (0)	2 (6)	0 (0)	0 (0)	0 (0)	0 (0)	1 (3)
PBulletin	82	59 (72)	22 (27)	21 (26)	16 (20)	23 (28)	0 (0)	13 (16)	6 (7)	9 (11)	30 (37)
RER	38	20 (53)	5 (13)	10 (26)	7 (18)	6 (16)	1 (3)	0 (0)	2 (5)	1 (3)	9 (24)
Lancet	10	6 (60)	0 (0)	0 (0)	2 (20)	1 (10)	1 (10)	4 (40)	1 (10)	2 (20)	3 (30)

^aMA = meta-analytic review; Emp. assessment of pb = empirical assessment of publication bias; Source comparison = subgroup comparison by source of the data; T&F = trim and fill; CMA = cumulative meta-analysis; Egger's test = Egger's test of the intercept; B&M's test = Begg and Mazumdar's rank correlation test; Other = other tests to assess publication bias; Multiple methods = multiple methods were used to assess publication bias. ^bAMJ = Academy of Management Journal; JAP = Journal of Applied Psychology; JOM = Journal of Management; PPsych = Personnel Psychology; PBulletin = Psychological Bulletin; RER = Review of Educational Research. ^cFunnel plot analyses were interpreted independent of other publication bias assessments (e.g., trim and fill and Egger's test). ^dTotal of AMJ, JAP, JOM, and PPsych.

presence and influence of publication bias on validity estimates. Of the more appropriate publication bias methods, we found only four studies in the management and I/O psychology journals that used the trim and fill technique (Duval & Tweedie, 2000a, 2000b). No study used a cumulative meta-analysis (Borenstein et al., 2009; McDaniel, 2009), Egger's test of the intercept (Egger, Davey Smith, Schneider, & Minder, 1997), or Begg and Mazumdar's (1994) rank correlation test. Finally, when examining the results for only 2009–2010, the conclusions are almost exactly the same as the entire review period from 2005 to 2010, suggesting that the state of affairs is not improving.

The review of the management and I/O psychology literature may be contrasted with the results from top-tier journals in psychology (*Psychological Bulletin*), education (*Review of Educational Research*), and medicine (*Lancet*). For instance, 72% of meta-analytic reviews in *Psychological Bulletin*, 53% of the reviews in *Review of Educational Research*, and 60% of the reviews in *Lancet* empirically examined the issue of publication bias (compared with 31% in management and I/O psychology). Many of these studies used multiple publication bias methods to evaluate the results as one can have greater confidence in the conclusions when multiple methods are in agreement. The benefit of using multiple methods is supported by the concept of *triangulation*, which is the use of 'multiple reference points to locate an object's exact position' (Jick, 1979, p. 602). In the social sciences, triangulation can be used to justify the use of multiple study designs, settings, samples, and methods to examine a relation of interest between two variables (Campbell & Fiske, 1959; Sackett & Larson, 1990). Therefore, it is advantageous to employ multiple methods in the evaluation of publication bias.

Finally, the results from these journals also suggest that more appropriate methods are used, although inadequate ones are also employed. This is an indication that efforts are being made to evaluate the threat of publication bias in a more comprehensive and rigorous manner. As a final comparison, the leading publisher of systematic reviews in the medical sciences, the *Cochrane Database of Systematic Reviews* (IF = 5.7), requires that all articles address the issue of publication bias (Higgins & Green, 2009). Furthermore, the use of the Failsafe N and related methods are not recommended (Higgins & Green, 2009).

Overall, the results of our review suggest that the majority of meta-analytic reviews in our field ignore the issue of publication bias. We do not argue that all meta-analytic reviews suffer from publication bias. However, given that we do not routinely evaluate our meta-analytic data for publication bias, we do not know the extent to which it distorts our literatures. One possible reason for the lack of attention and concern with the issue of publication bias is the absence of methodologi-

cal reviews on the topic in our peer-reviewed journals (we are unaware of any reviews other than McDaniel et al., 2006). For instance, a key word search using 'publication bias' in our leading research methods journal, *Organizational Research Methods*, identifies just a handful of articles which simply mention the issue in passing. Yet the issue of publication bias is of increasing importance. The American Psychological Association (2010) style manual now recommends an assessment of publication bias in all meta-analytic reviews as a sensitivity analysis. Thus, the current paper presents an important contribution to scholarship in light of the possible effects of publication bias.

1.1. The current study

We first review the issue of publication bias and several of its likely causes. Next, we review advanced methods for the evaluation of publication bias. We then apply these methods using a previously published meta-analytic review by Berry, Sackett, and Tobares (2010) on conditional reasoning tests of aggression (CRT-A). As the meta-analytic review by Berry et al. (2010) did not provide such an analysis, the current paper complements Berry et al.'s meta-analysis by filling this gap. Thus, we conduct the analyses using the same dataset as Berry et al. (2010).³

A primary reason for the assessment of publication bias in the CRT-A literature relates to its recency, which allows one to examine the potential presence of the time-lag bias as a potential cause for publication bias. The time-lag bias has been found in various literatures in the medical sciences (e.g., Ioannidis, 2005, 2006; Trikalinos & Ioannidis, 2005; Uitterlinden et al., 2006), and has not been studied or assessed in the management and I/O psychology literature. The time-lag bias occurs if earlier effect sizes (e.g., correlation coefficients) are larger than effect sizes obtained in later time periods (Ioannidis, 2005; Trikalinos & Ioannidis, 2005). Thus, the time to publication is shorter for statistically significant effects than for statistically insignificant effects (Ioannidis, 1998, 2005; Stern & Simes, 1997; Trikalinos & Ioannidis, 2005). The time-lag bias could also include the Proteus effect (i.e., studies with large effects are published earlier because they are more dramatic and more interesting; Trikalinos & Ioannidis, 2005).

Under either explanation, validity studies in relatively new literatures may be subject to a bias, such that initial findings overestimate the validity of a test. Furthermore, this type of bias is likely to occur in 'interesting' areas where research productivity is rapid and relies on non-randomized designs (Trikalinos & Ioannidis, 2005). Research on CRT-A scales is both relatively new and quite interesting, thus making the presence of a time-lag bias possible.

1.2. Causes of publication bias

It is likely that publication bias has existed since scientific research was first conducted and reported. However, it was only with the establishment of systematic reviews and meta-analysis that the issue has gained an importance (Rothstein et al., 2005b). Additionally, it was not until the development of improved methods that researchers were able to begin to identify empirically the presence of publication bias and to assess the extent to which it can bias the direction and magnitude of parameter estimates.

It has been proposed that publication bias is present in all areas of research, but the extent to which this bias is present is likely to vary across literature areas (Sutton, 2005). Publication bias is the result of information suppression mechanisms that transpire from the moment a study is initiated to the point its results are disseminated to consumers of research (Rothstein et al., 2005b). In addition to time-lag bias, various causes of publication bias exist. Two of the more frequently discussed causes are editor and reviewer decisions (Dickersin, 2005). In these instances, studies are rejected in the editorial review process because the sample size was small, the results were not statistically significant (Greenwald, 1975), the results were contrary to theory, contrary to the position of the editor/reviewers, or contrary to purportedly well-established knowledge (Banks & McDaniel, 2011; Davis, 1971; Dickersin, 2005).

However, publication bias can also occur because of author decisions. In some fields, such decisions are likely the primary cause (Dickersin, 2005), partly because they typically occur prior to editor/reviewer decisions in the research dissemination process and because authors have more control over their data. For instance, an author may never submit a study to a journal or conference because the study had a small sample size, the results were statistically insignificant, contrary to theory, contrary to trends of past research, or contrary to the position of the author (e.g., an employee of a consulting company does not report that a structured interview they developed for a client did not work). In the event an author submits a paper, the author may decide to not report certain results to save space, or because the findings are not interesting, contrary to theory, or contrary to their position on a topic (e.g., an author does not report that a situational judgment test has poor validity because he or she believes such tests to have high validity; Banks & McDaniel, 2011). Editors and reviewers can also request that results be removed during the editorial process. If a study ultimately remains unpublished, primary researchers may be noncompliant with requests for unpublished samples from meta-analytic researchers for a variety of reasons, such as disinterest and time constraints or because unpublished studies had been already discarded (Ferguson & Brannick, 2012).

Organizational constraints are another cause of publication bias. For instance, results may not be submitted because the data are proprietary in nature and the release could lead to litigation issues (e.g., a structured interview that has both zero validity and adverse impact). It is also possible that organizational researchers with monetary interests in a product (e.g., a commercially marketed employment test or emotional intelligence training program) may suppress a study because the results would damage sales of the product (Banks & McDaniel, 2011; McDaniel et al., 2006).

Another cause of publication bias concerns accessing grey literature, such as conference papers, dissertations, and technical reports (Banks & McDaniel, 2011; Hopewell et al., 2005; Rothstein & Hopewell, 2009). For example, many conference papers do not get published in journals and, unfortunately, are not always distributed to those who request a copy. It is also possible that a paper is not identified in a systematic literature search, or it is impossible to be located or retrieved (this sometimes occurs in an inter-library loan program). Furthermore, it can occur that researchers conducting a systematic review elect not to translate foreign-language articles (Banks & McDaniel, 2011; Hopewell et al., 2005).

Finally, although the classic case of publication bias is an overestimation of the parameter estimate, publication bias can result in the underestimation of effect size estimates. For example, researchers may be more likely to report mean racial differences when they are small than when they are large. For instance, McKay and McDaniel (2006) found smaller mean racial differences in job performance in published studies than in technical reports. This may indicate that the published literature is underestimating mean racial differences in the population (see also Tate & McDaniel, 2008). Thus, the direction of publication bias may vary depending on the research topic and the mechanism that causes the bias.

1.3. Arguments against publication bias in the social sciences

It has been previously suggested that publication bias may not be a significant threat to meta-analytic results (e.g., Hunter & Schmidt, 2004, pp. 496–498). First, primary studies in the social sciences (e.g., education, general psychology, organizational sciences) typically report multiple hypotheses and associated results. Thus, whether a study is published may not depend on the statistical significance of any one result. Second, meta-analytic studies in the social sciences may be based on results that are not the focal issue of primary studies, and therefore, the extent to which those results are disseminated may not depend on whether or not those results are statistically significant.

However, the testing of multiple hypotheses may mitigate but not necessarily eliminate the potential for

publication bias. For instance, reporting preferences may still be given to statistically significant outcomes or subgroups (Sutton, 2009) as well as to results that are socially pleasant as is the case when studying mean racial differences (e.g., report differences when small but not when large; Tate & McDaniel, 2008). Thus, selective reporting may still occur. Also, although meta-analytic reviews may be based on results not of focal interest by primary studies, there are a large number of meta-analytic reviews that are based on the exact same focal topic as primary studies. Research has indicated that publication bias can occur within individual literature areas when researchers do not report results from subgroups (Tate & McDaniel, 2008) or studies with results counter to prevailing theory have not been submitted for publication (e.g., Banks, Batchelor, & McDaniel, 2010). Thus, the assertion that publication bias is not a problem in the organizational sciences lacks empirical support. Furthermore, even if publication bias is not a problem for the field at large, it may still occur within individual research areas.

Publication bias likely exists to varying degrees in the social sciences (e.g., education, general psychology, organizational sciences, etc.). Thus, it is important that each field and each individual literature area attempts to evaluate publication bias as a potential threat to its meta-analytic results. The results presented in our Table 1 indicate that the organizational sciences lag behind other areas of research in considering publication bias. We hope that the following study assists in closing this gap.

2. Method

2.1. Data source

The data used in this study were provided by Berry et al. (2010). Over the last decade, promising work has been conducted by James and colleagues (e.g., James, 1998; James & Mazerolle, 2002; James et al., 2005) in the development of a new personnel selection measure, known as CRT-A. This line of research is an attempt to develop a measure that makes it more difficult for applicants to fake and to explain more variance in work outcomes than conscious self-reports. Recently, the meta-analytic review by Berry et al. (2010) offered evidence of validity for these measures, but with validities of lower magnitude than offered by James et al. (2005).

2.2. Meta-analysis procedure

There are many methods for evaluating publication bias. Some methods are more appropriate than others. For example, the trim and fill analysis (Duval & Tweedie, 2000a, 2000b) has demonstrated clear advantages over

the failsafe N (Aguinis et al., 2011b; Becker, 2005; Geykens et al., 2009; McDaniel et al., 2006). The use of cumulative meta-analysis has also been offered as a useful tool for the detection of publication bias (Borenstein et al., 2009; McDaniel, 2009). Although research has not yet fully evaluated which publication bias methods are the most optimal under different conditions, what is clear is that multiple methods for evaluating publication bias are recommended in identifying the presence and influence of publication bias. To the extent that multiple methods are consistent with an inference of publication bias, one can have greater confidence in the conclusion of publication bias.

Multiple publication bias methods were used in this study. Comprehensive meta-analysis (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005) was used to complete the following analyses: (a) a meta-analysis of observed correlations using both random-effects and fixed-effects models; (b) a trim and fill analysis on random and fixed-effects models; (c) a cumulative meta-analysis with correlations sorted by precision (i.e., the inverse of the standard error of the correlation) from high to low and a cumulative meta-analysis with correlations sorted by publication date from most distant to most recent; (d) Egger's test of the intercept (Egger et al., 1997); (e) a subgroup analysis; and (f) a one-sample removed sensitivity analysis. A review of these methods is presented next.

The CMA software implements meta-analysis consistent with the Hedges and Olkin tradition of meta-analysis (Borenstein et al., 2009; Hedges & Olkin, 1985). When compared with psychometric meta-analysis software (e.g., Schmidt & Le, 2005), CMA differs in two ways. First, CMA assigns different weights to the correlations. For analysis of fixed-effects models, CMA uses weights derived from sample size, that is, it uses the inverse of the sampling error variance. In its analysis of random-effects models, CMA modifies this weight to also reflect the variance that is not attributable to sampling error. Psychometric meta-analysis is always a random-effects model, but uses sample size as the study weight in a bare bones analysis (an analysis in which random sampling error is the only artifact considered). In this study, the correlation between sample size and the precision weight in the CMA fixed-model analysis was .95. For the CMA random-effects analysis, the correlation between sample size and the random-effects weights was .77. As a result, the findings of CMA's fixed-effects model are virtually identical to the results of a bare bones random-effects psychometric meta-analysis (both analyses yield a mean correlation of .16 for counterproductive work behavior (CWB) criteria in this study; see also Berry et al., 2010). Due to CMA's weights used in random-effects model analyses, effects from larger sample size studies are given less relative weight than in random-effects model analyses in psychometric meta-analysis.

The second way that the meta-analytic approaches differ is that psychometric meta-analysis conducts analyses of correlations, but CMA converts the correlations into the Fisher z metric, conducts the analyses with the Fisher z , and then converts the results back into correlations. There is an ongoing debate in the meta-analytic literature concerning the merits of Fisher z transformations (e.g., Field, 2001, 2005; Hafdahl, 2009, 2010), but that difference in results between methods is usually minimal (e.g., Hafdahl, 2010; Hafdahl & Williams, 2009). In summary, differences in observed mean validities between Berry et al. (2010) and this study are attributable to differences in the meta-analytic approach used. These differences do not affect the conclusions concerning the evidence of publication bias.

2.2.1. *Trim and fill*

The trim and fill analysis (Duval & Tweedie, 2000a, 2000b) is a popular analysis for detecting publication bias and estimating the extent to which publication bias distorts the mean correlation (Geyskens et al., 2009). The technique uses a funnel plot (Light & Pillemer, 1984) that graphically displays the distribution of effect sizes from a meta-analytic study. In the traditional funnel plot, study correlations expressed as Fisher z are plotted on the X-axis (ordered from right to left, largest to smallest) and study precision is plotted along the Y-axis (ordered from top to bottom, most precise to least precise; Sterne & Egger, 2005).

Because the standard error is a function of sample size and the magnitude of an effect, precision is highly correlated with sample size (.95 in our analysis using the fixed-effects model; .77 using the random-effects model). Large sample studies yield more precise correlations (i.e., the correlations have smaller standard errors). For this reason, large sample studies typically cluster together around the center line of a funnel plot (and are high on the Y-axis). Small sample studies yield less precise correlations and can, to a greater extent than correlations from larger samples, over- or underestimate the magnitude of validity. As a result, small sample studies commonly scatter widely across the X-axis (and are low on the Y-axis). In the event that all sample correlations are graphed, and assuming that random sampling error is the only operating artifact, the distribution of samples will be symmetrical (Duval, 2005). If small sample studies or nonsignificant results are missing from a funnel plot (typically missing to the left; McDaniel et al., 2006), asymmetry is present and is interpreted as an indication of potential publication bias (Duval, 2005).

Trim and fill also evaluates the degree to which an effect size is affected by publication bias. The method imputes samples (i.e., correlations) that would be needed to make an asymmetrical distribution symmetrical, and then adjusts the observed mean correlation to

account for the imputed samples. The technique proceeds in an iterative process that 'trims' extreme samples from the asymmetric distribution until the distribution is symmetrical. In the last step, the funnel plot is 'filled' by adding the trimmed samples back to the graph with an imputed study to offset each. Trim and fill then recalculates the mean effect size of the symmetrical distribution (i.e., the 'trim and fill adjusted mean effect size'). One can then compare the difference between the observed mean effect size and the adjusted mean effect size.

Thus, the trim and fill analysis not only detects the potential presence of publication bias, but also provides an estimate of its amount. For instance, if the difference between the observed and adjusted effect size is small in magnitude, one should interpret the effects of publication bias as minimal. If the difference between the observed effect and the adjusted effect is large, but the conclusion remains unchanged that X is correlated with Y , publication bias is moderate. Finally, if the conclusion changes as to whether the relationship between X and Y is of a meaningful magnitude, one can interpret publication bias as severe (McDaniel et al., 2006; Rothstein et al., 2005b).

The primary limitation of the trim and fill method is that it is based on the assumption that random sampling error is the only source of variance across samples. If a population is heterogeneous, the trim and fill method may not derive accurate estimations (Terrin, Schmid, Lau, & Olkin, 2003). For example, if a distribution of samples is bimodal (e.g., the correlation magnitude differs by sex), the analysis is less likely to be able to correctly impute missing studies. In the event that theory and statistics indicate the presence of moderators, publication bias analyses on sample subgroups should be conducted.

2.2.2. *Cumulative meta-analysis*

A cumulative meta-analysis is conducted by sorting effect sizes (e.g., a correlation) by a characteristic of interest, adding one effect size at a time to the meta-analysis, and recalculating the analysis with each additional effect size. The classic example of a cumulative meta-analysis is the examination of medical studies sorted in chronological order to identify the time point when the cumulative effect size stabilizes (e.g., Lau & Antman, 1992; Lau, Schmid, & Chalmers, 1995). In the context of publication bias, samples can be sorted by their precision from low to high or by publication date. In a cumulative meta-analysis, the cumulative means can be plotted in a forest plot and examined for evidence of 'drift' as samples are added to the meta-analysis. In the event publication bias is present, the cumulative effect size will drift to one side of the forest plot as samples are added. In a typical example of publication bias, where small magnitude and small sample studies are absent

from a distribution of studies (Dickersin, 2005), the effect size, sorted by precision, will drift in a positive direction (McDaniel, 2009).

The cumulative meta-analysis can also be used to evaluate a time-lag bias as a potential cause of publication bias. In the case of a time-lag bias, the effect sizes, sorted by time in a cumulative meta-analysis, will drift in a negative direction as smaller magnitude more recent effect sizes are added to the analysis (Ioannidis, 1998; Stern & Simes, 1997; Trikalinos & Ioannidis, 2005). Thus, a drift in a cumulative meta-analysis can provide an indication of the existence of publication bias. The difference between the cumulative effect sizes that demonstrate stability and those that demonstrate drift can be used to estimate the magnitude of the influence of publication bias.

2.2.3. Egger's test of the intercept

Egger's test of the intercept (Egger et al., 1997) is one of several regression-based methods for the assessment of publication bias (Sterne & Egger, 2005). This test proposes that publication bias can be detected by predicting a standardized effect with precision. The effect size is reported as the slope of a regression line and publication bias is reported as the intercept. It has been suggested that this method has more power than Begg and Mazumdar's (1994) rank correlation, another regression-based method to detect publication bias (Sterne & Egger, 2005), partly because it better accounts for factors that may affect the detection of publication bias (e.g., sample size and effect size; Borenstein et al., 2009). We thus use Egger's test of the intercept and not Begg and Mazum-

dar's rank correlation test as both are conceptually similar (Sterne & Egger, 2005), but Egger's test of the intercept has clear advantages (Borenstein et al., 2009). However, the test does not provide an estimate of the degree to which an effect size may be over- or underestimated (as trim and fill does), nor does it depict any trend to assess publication bias over precision or time (as the cumulative meta-analysis does).

2.2.4. Additional analyses

Four subgroup analyses were run to test the likelihood of publication bias in subgroups (e.g., version of test, student vs. nonstudent samples, dichotomous vs. nondichotomous criterion, and published vs. grey literature). Furthermore, a one-sample removed analysis and an outlier analysis were conducted to test the robustness of the publication bias analyses (Borenstein et al., 2009; Beal, Corey, & Dunlap, 2002). The one-sample removed analysis computes the effect size repeatedly by removing one sample in each iteration (with replacement). The result of the sensitivity analysis examines the robustness of the effect size and its potential range if any single sample was removed.

3. Results

Table 2 displays the samples included in our analyses. There were 21 unique samples with a total of 3,820 individuals. The first two columns display the study ID and the author(s) of the study. The next three columns

Table 2. Summary of samples included in the meta-analysis

Study ID	Author(s)	Sample size	Observed correlation	Criterion
1.	James and McIntyre (2000)	105	.34	CWB
2.	James and McIntyre (2000)	111	.43	CWB
3.	Hawes (2000)	349	.06	CWB
4.	Bing et al. (2007)	176	.07	CWB
5.	Patton (1998)	100	.38	CWB
6.	James and McIntyre (2000)	135	.31	CWB
7.	Russell, James, and McIntyre (2004)	191	.40	CWB
8.	Frost (2005)	183	.40	CWB
9.	Frost (2002)	191	.25	CWB
10.	Walton (2004)	770	-.06	CWB
11.	Lebreton (2002)	105	-.11	CWB
12.	Lebreton (2002)	121	-.10	CWB
13.	Lebreton (2002)	130	.13	CWB
14.	Sablynski and Mitchell (2006)	95	.28	CWB
15.	James and McIntyre (2000)	188	.37	CWB
16.	Bing et al. (2007)	225	.22	CWB
17.	Bing et al. (2007)	62	.31	CWB
18.	Hawes (2000)	395	.12	Job performance
19.	Hornick et al. (1999)	68	-.03	Job performance
20.	Hornick et al. (1999)	52	.31	Job performance
21.	Hornick et al. (1999)	68	.31	Job performance

CWB = counterproductive work behaviors.

display the sample size for each sample, the observed correlation, and the criterion type (e.g., CWB or job performance).

Table 3 displays the summary of the meta-analysis of the observed correlations as well as the publication bias analyses (i.e., trim and fill analysis and Egger's test of the intercept) by criterion. The first two columns indicate the type of criterion (i.e., overall performance as well as CWB and job performance separately) and the number of observed samples (*k*) included in the analysis. The next four columns report the mean observed correlation (\bar{r}_{obs}) and the confidence interval (95% CI) of the meta-analysis by fixed-effects and random-effects meta-analytic models. The next seven columns display the results of the trim and fill analysis by fixed-effects and random-effects models; the number of imputed correlations (*ik*) to achieve symmetry in the funnel plot, the trim and fill adjusted observed mean (adj. \bar{r}_{obs}), the adjusted CI (adj. 95% CI), and the difference between the observed mean correlation and the trim and fill adjusted mean correlation ($\Delta\bar{r}_{obs}$). The last column displays the results of Egger's test of the intercept (Egger's test).

Table 3 contains results for both fixed- and random-effects meta-analysis models to demonstrate the minor differences between both techniques (the conclusions of the analyses do not change). We focus on the results from the random-effects models, and discuss their implications as such models provide more accurate estimates given our data and research inquiry (Borenstein et al., 2009; Hunter & Schmidt, 2004). The results displayed in the table are consistent with an inference of publication bias. For instance, in the random-effects analysis using CWB as the sole criterion, seven correlations were imputed to create a symmetrical distribution (see Figure 1). The mean observed correlation (.22) was adjusted downward to .08, resulting in a difference of .14. Similarly, the observed 95% CI (.12 to .31) was

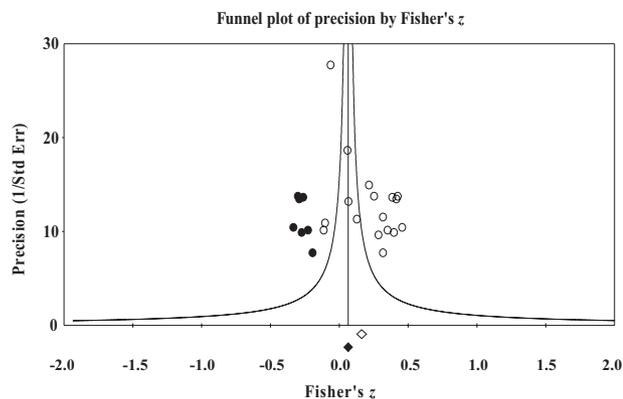


Figure 1. Funnel plot of CWB observed correlations (*k* = 17) with trim and fill imputation. Imputed correlations are indicated in black (*ik* = 7).

Table 3. Meta-analytically derived validity coefficients and publication bias analysis^a

Criterion	Meta-analysis results				Trim and fill results ^c				Random effects				Egger's test int (p value)
	Fixed-effects		Random-effects		Fixed effects		Random effects		Random effects		Random effects		
	<i>k</i>	\bar{r}_{obs}	95% CI	\bar{r}_{obs}	95% CI	<i>ik</i>	adj. \bar{r}_{obs}	adj. 95% CI	$\Delta\bar{r}_{obs}$	adj. \bar{r}_{obs}	adj. 95% CI	$\Delta\bar{r}_{obs}$	
Overall perf. ^b	21	.16	.13 to .19	.21	.13 to .29	8	.08	.05 to .10	.08	.09	.00 to .18	.12	3.66 (.01)
CWB	17	.16	.13 to .19	.22	.12 to .31	7	.06	.03 to .10	.10	.08	-.02 to .19	.14	4.83 (.01)
Job perf.	4	.14	.06 to .22	.16	.02 to .30	1	.12	.04 to .20	.02	.13	-.01 to .26	.03	1.12 (.30)

^a*k* = number of samples (i.e., number of correlation coefficients); \bar{r}_{obs} = weighted mean observed correlation; 95% CI = 95% confidence interval; *ik* = number of trim and fill imputed correlations; adj. \bar{r}_{obs} = trim and fill adjusted observed mean (the weighted mean of the distribution of the combined observed correlations and the imputed correlations); adj. 95% CI = trim and fill adjusted 95% confidence interval; $\Delta\bar{r}_{obs}$ = difference between \bar{r}_{obs} and adj. \bar{r}_{obs} ; Egger's test = Egger's test of the intercept. ^bOverall perf. = CWB and job perf. (i.e., entire sample). ^cWe offer that a mean effect size estimate that results from publication bias analyses, such as trim and fill, should be referred to as an 'adjusted estimate'. We discourage the use of the phrase 'correcting for publication bias' so as not to confuse the adjusted estimate with a population parameter estimated by using psychometric corrections (e.g., measurement and range restriction corrections, Hunter & Schmidt, 2004). CWB = counterproductive work behavior.

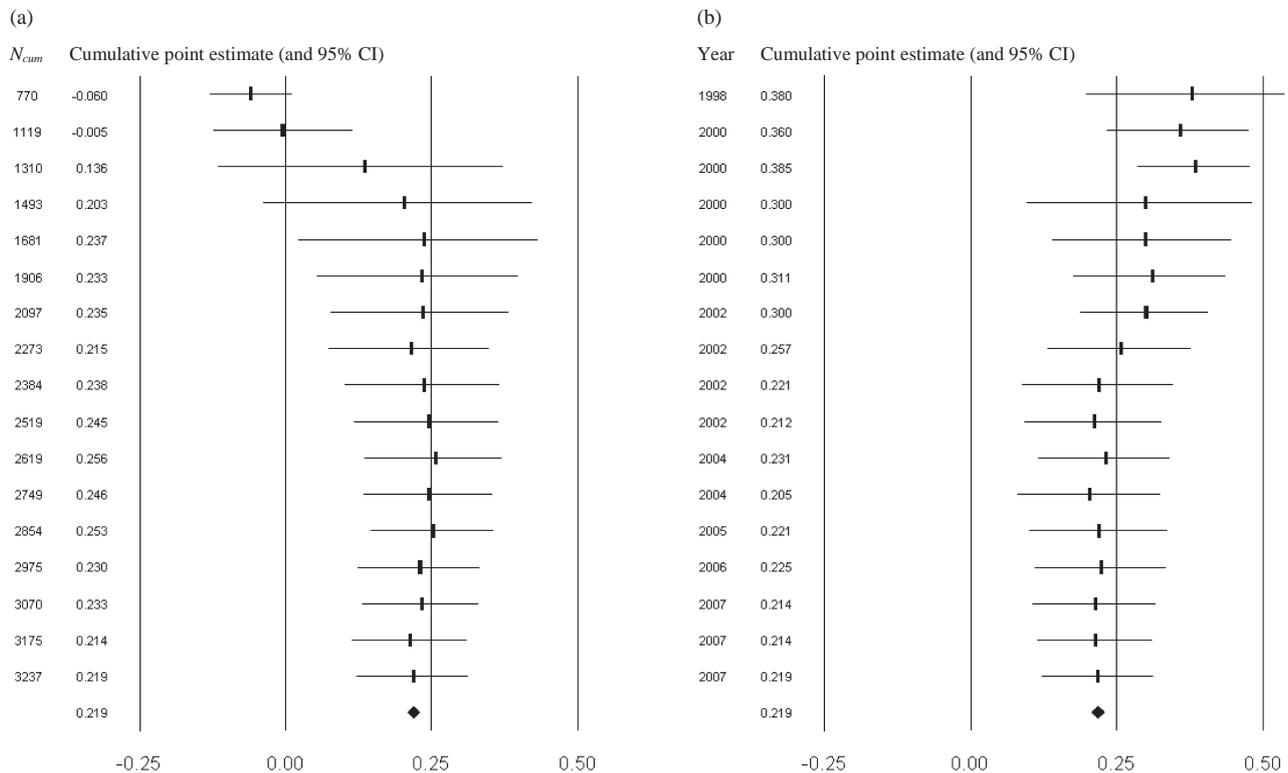


Figure 2. Forest plots of the cumulative meta-analyses for CWB. (a) Cumulative meta-analysis by precision. The sample sizes and the point estimate in each row are cumulative (i.e., N_{cum} = cumulative sample size). (b) Cumulative meta-analysis by publication year. The point estimate in each row is cumulative; year = publication year.

adjusted downward ($-.02$ to $.19$) and includes zero. The difference in observed means ($\Delta\bar{r}_{obs}$) is judged to be substantial ($.14$; a difference of 64%), which is consistent with an inference of publication bias and the conclusion that small magnitude correlations may likely be missing. Egger's test supports this conclusion ($4.83, p < .01$).

3.1. Cumulative meta-analyses

Figure 2 displays the forest plots of our cumulative meta-analyses of the CWB samples (the forest plot of all samples is available from the first author); the cumulative meta-analyses by precision (panel (a)) and by publication year (panel (b)). In both instances, the respective meta-analysis is recomputed as samples are added one by one. For the cumulative meta-analysis by precision, the most precise samples (i.e., the largest samples) are added first and the least precise samples (i.e., the smallest samples) are added last. The forest plot (see Figure 2, panel (a)) indicates a clear positive drift. The cumulative point estimate of the first two samples is $-.01$ (cumulative $N [N_{cum}] = 1,119$; 35% of the sample size summed across all samples). As less precise samples are added, the cumulative point estimate drifts positively such that by the time the cumulative sample size reaches 53% ($N_{cum} = 1,888$) of the total sample size across samples, the point esti-

mate is $.24$. After that, the point estimate stabilizes at around $.22$, which is also the final estimate. The overall drift from $-.01$ (35% of N) to $.22$ (100% of N) seems dramatic ($\Delta = .23$). Consistent with the results from the trim and fill analysis, the cumulative meta-analysis by precision is consistent with the inference that the data on CRT-A tests are adversely affected by publication bias such that it overestimates the validity of CRT-A tests.

To test for the presence of a time-lag effect, we sorted the samples by the year they were published, and conducted a second cumulative meta-analysis by year of publication. Consistent with the time-lag effect, including the Proteus effect, the forest plot (see Figure 2, panel (b)) shows a clear negative drift, suggesting that samples with large effects are published earlier, potentially because they are more dramatic and more interesting, or because small effect-size samples take substantially longer to appear in print (Ioannidis, 1998; Stern & Simes, 1997; Trikalinos & Ioannidis, 2005). In conclusion, several separate analyses yield results consistent with an inference of publication bias.

3.2. Additional analyses

We conducted additional analyses to examine the robustness of our results. Specifically, we examined the

Table 4. Meta-analytically derived CWB validity coefficients and publication bias analysis by subgroup^a

	Meta-analysis			Trim and fill results ^b				Egger's test
	<i>k</i>	\bar{r}_{obs}	95% CI	<i>ik</i>	adj. \bar{r}_{obs}	adj. 95% CI	$\Delta\bar{r}_{obs}$	int (<i>p</i> -value)
Version of test ^c								
CRT-A	12	.17	.06 to .29	4	.05	-.09 to .18	.12	4.61 (.03)
Used student sample								
Yes	11	.19	.06 to .32	5	.04	-.10 to .18	.15	4.04 (.06)
No	6	.26	.12 to .39	3	.13	-.03 to .27	.13	7.60 (.01)
Used dichotomous criterion								
No	9	.30	.19 to .40	1	.28	.18 to .37	.02	7.10 (.03)
Yes	7	0.09	-.04 to .23	2	.01	-.13 to .15	.08	3.01 (.11)
Source type								
Published literature	7	.29	.19 to .38	2	.24	.13 to .33	.05	3.06 (.19)
Gray literature	10	.17	.03 to .30	3	.05	-.10 to .19	.12	4.32 (.06)

^a*k* = number of samples (i.e., number of *d* coefficients); \bar{r}_{obs} = weighted mean observed correlation; 95% CI = 95% confidence interval; *ik* = number of trim and fill imputed correlations; adj. \bar{r}_{obs} = trim and fill adjusted observed mean (the weighted mean of the distribution of the combined observed correlation and the imputed correlation); adj. 95% CI = trim and fill adjusted 95% confidence interval; $\Delta\bar{r}_{obs}$ = difference between \bar{r}_{obs} and Adj. \bar{r}_{adj} ; Egger's test = Egger's test of the intercept. ^bThe number of studies imputed is the same for both the fixed-effects and random-effects models because a fixed-effects model is the optimal technique to trim and fill studies from the funnel plot in the iterative process (Duval, 2005; Moreno et al., 2009; Sutton, 2005; Terrin et al., 2003). ^cThe number of samples was only large enough to conduct publication bias analyses on the CRT-A version of the test. The other two versions (e.g., CRT-developmental and VCRT) had a *k* of three and two, respectively. CRT-A = conditional reasoning tests of aggression; CWB = counterproductive work behavior.

results for subgroups of our data. In addition, we replicated our analyses by dropping one sample at a time and repeating the analyses without the one sample dropped (i.e., one-sample removed analysis; Borenstein et al., 2009). We also used Beal et al.'s (2002) procedure to detect possible outliers. To the extent that these additional analyses support our prior results, one can place greater confidence in the conclusions drawn.

3.2.1. Subgroup analyses of CWB

Four subgroup analyses were conducted using the CWB samples. Berry et al. (2010) identified three primary subgroups: version of test, student or nonstudent sample, and dichotomous or nondichotomous criterion. We added a fourth subgroup analysis, a comparison of published literature (e.g., journals and test manuals) to grey literature (e.g., dissertations and conference papers; Hopewell et al., 2005; Rothstein & Hopewell, 2009). The results of these analyses are presented in Table 3. Cumulative meta-analyses of these subgroups are available from the first author. The subgroup analyses are limited due to the number of samples used (i.e., it is possible that not all samples have been detected). The results are supportive of an inference of publication bias. We note a difference of .12 between the published and grey literature. In summary, the results provide evidence consistent with an inference that the validity of the CRT-A scales might be overestimated due to publication bias (Table 4).

3.2.2. One-sample removed

We conducted a one-sample removed analysis to evaluate the sensitivity of our results (Borenstein et al., 2009).

Specifically, we conducted the analyses multiple times, each time deleting a different correlation. The results are virtually identical for all analyses (i.e., trim and fill analyses, cumulative meta-analysis, and Egger's test of the intercept). In only 1 out of 17 incidents did our results change substantially and that was when we removed sample 10 (*N* = 770; Walton, 2004). A comprehensive outlier analysis using Huffcutt and Arthur's (1995) sample adjusted meta-analytic deviancy statistic with corrections advocated in Beal et al. (2002) supported the results from our one-sample removed analysis.

However, the CRT-A scale used in the Walton sample is identical to most of the other scales; the observed correlation (–.06) is not an apparent outlier as several studies reported negative correlations (e.g., Hornick, Fox, Axton, & Wyatt, 1999; LeBreton, 2002; Walton, 2004), two with correlations at or below –.10 (LeBreton, 2002). The only difference is its size (*N* = 770), which is substantially larger than the other samples. This gives the sample more weight in our analyses because it is the most precise estimate of the population parameter (Borenstein et al., 2009; Hunter & Schmidt, 2004), making us reluctant to advocate its deletion. In addition, because this sample was part of Berry et al.'s (2010) data set, we saw no reason to judge the results without this sample as more correct.

4. Discussion

The majority of management and I/O psychology meta-analytic reviews do not address the issue of publication bias. This is inconsistent with other leading journals in the field of psychology (e.g., *Psychological Bulletin*), educa-

tion (e.g., *Review of Educational Research*), and medicine (e.g., *Lancet* and the *Cochrane Database of Systematic Reviews*). Of the meta-analytic reviews in management and I/O psychology that did empirically evaluate publication bias, the majority used inadequate methods. After a review of the issue of publication bias and the methods that can be used to evaluate it, we used several methods to assess whether publication bias has affected the validity estimates of CRT-A scales.

Overall, results are consistent with an inference of publication bias such that the results of Berry et al. (2010) overestimated the validity of the CRT-A scales. For the CWB distribution, the trim and fill analysis estimated an adjusted observed mean (.08) of less than half of the unadjusted mean observed correlation (.22). In addition, the observed CI (.12 to .31) was adjusted down (-.02 to .19) and included zero. The difference in means ($\Delta = .14$) is consistent with an inference of publication bias such that the results identified by Berry et al. (2010) may be overestimating the validity of the CRT-A. Furthermore, the differences could be judged severe in that practitioners might decide to use the test if the validity were .22 for CWB, but few might decide to use the test if the validity were .08 (Rothstein et al., 2005b).

The cumulative meta-analyses provided additional evidence consistent with an inference of publication bias. When sorting by precision, the forest plot showed a clear positive drift, consistent with the inference that small samples with small magnitude effect sizes are missing from the available literature. In addition, when sorting by publication date, a waning temporal pattern in the parameter estimate emerged, suggesting the presence of the time-lag bias (e.g., the Proteus effect; Trikalinos & Ioannidis, 2005). Furthermore, Egger's test of the intercept also provided evidence consistent with an inference of publication bias. Finally, the various subgroup analyses, particularly the version of test, student and nonstudent samples, and the comparison of grey literature to published literature provided evidence consistent with an inference of publication bias.

4.1. Limitations

We shared an earlier version of this paper with several scholars with stakes in the debate concerning the validity of the CRT-A scales. One scholar mentioned concerns regarding our trim and fill analysis. A meta-analysis may be conducted with either a fixed-effects model or random-effects model. A random-effects model is preferred (Borenstein et al., 2009; Hunter & Schmidt, 2004). Similarly, in a trim and fill analysis, when trimming and imputing effect sizes to achieve symmetry in a funnel plot, a fixed-effects or random-effects model can be used. Thus, in addition to a decision regarding which meta-analysis model to employ (fixed or random), there

is also a decision concerning whether the trim and fill analysis should be conducted with a fixed- or random-effects model. For the trim and fill analysis, the fixed-effects model is preferred even though the random-effects model should be used for all other meta-analytic procedures (Sutton, 2005) as fixed-effects models (a) are less affected by funnel asymmetry (e.g., the random-effects trim and fill incorrectly adjusts for samples that are not necessarily suppressed; Terrin et al., 2003) and (b) give less weight to less precise samples than random-effects models (Sutton, 2005, 2009).

In addition, there is also a decision concerning which estimator of the number of missing studies (*L*- or *R*-estimator) to use. When conducting the trim and fill analysis, we used the fixed-effects model with the *L*-estimator. This estimator is the preferred and most often used approach (Moreno et al., 2009; Sutton, 2005; Terrin et al., 2003), largely because it is more robust than the *R*-estimator, especially when the number of samples (*k*) is small (Duval, 2005; Sutton, 2005). Thus, we believe that the trim and fill method we used is the most appropriate analysis method available. Still, we acknowledge that fewer studies are imputed for the trim and fill procedure when a random-effects model is used to achieve symmetry and when the *R*-estimator is used. However, given the arguments presented above, we believe that our analysis method was optimal. We also note that inferences related to publication bias in these data rest on multiple lines of evidence, and not solely on our trim and fill analysis.

There are three additional potential limitations. First, there is the possibility that unknown heterogeneity resulted in inaccuracies in the trim and fill results. This potential limitation exists for any application of this method and thus is not a concern unique to this study. Furthermore, other methods support the results of the trim and fill analysis (e.g., cumulative meta-analysis and Egger's test). Also, the subgroup analyses, which removed potential heterogeneity (Hunter & Schmidt, 2004), provide evidence of publication bias. Second, there are only a limited number of publicly available samples using the CRT-A scales, partly because research on CRT-A scales is relatively new. Therefore, the conclusions reached in this study are restricted to the studies summarized by Berry et al. (2010). Future research will be needed for more credible estimates of the validity of CRT-A scales and publication bias analyses should be repeated as additional data become available.

Finally, the publication bias methods illustrated in this review were derived largely from the medical sciences and therefore, the methods were designed for observed correlations and do not consider the attenuating influence of measurement error and range restriction. Thus, not addressing the influence of measurement error is a potential limitation of these methods and it is subsequently an important avenue for future research. Conse-

quently, we recommend additional research (e.g., simulation studies) to consider the value of these methods when effect sizes are corrected for measurement error and/or range restriction.

4.2. Recommendations for future meta-analytic reviews

Given our results, three specific recommendations for future research emerge. First, we recommend that all meta-analytic reviews use multiple publication bias assessment methods as a sensitivity analysis to evaluate the extent to which the meta-analytically estimated effect sizes are robust to threats of publication bias. Furthermore, we encourage journals to publish reevaluations of previous meta-analytic reviews, regardless of whether or not publication bias was found to be a problem. This step is necessary to avoid 'publication bias in publication bias results' in which only meta-analytic reviews that find publication bias get published; giving the indication publication bias is a rampant problem (which it might not be).

Second, although publication bias methods have improved in their ability to detect and evaluate the magnitude of publication bias, prevention is still the best solution (Sutton, 2009). A recent review of trends in design, measurement, and analysis techniques published in our field called for more attention to be paid to issues of design (Aguinis, Pierce, Bosco, & Muslin, 2009). Consistent with this line of thinking, we recommend that organizations within the fields of management and I/O psychology (e.g., the *Society for Industrial and Organizational Psychology* and the *Academy of Management*) create research registries (Banks & McDaniel, 2011). Such registries, even if limited in capacity, could be useful in limiting the presence of publication bias in meta-analytic reviews (Berlin & Ghersi, 2005). Systematic searches could then include research registries in order to add additional published and unpublished samples to meta-analytic reviews. Although research registries would not completely eliminate the threat of publication bias, even an incomplete registry will serve to reduce the existence of publication bias.

Finally, journals in management and I/O psychology could provide supplemental information from primary and meta-analytic studies on their websites. This practice is applied in medical research (Evangelou, Trikalinos, & Ioannidis, 2005). Given page constraints, many journal articles in our field can only provide limited supplementary analyses, and meta-analytic reviews cannot always list effect size information in their study as is recommended (Hunter & Schmidt, 2004, p. 472). Thus, we recommend that journals add supplementary information to their websites in an attempt to help prevent publication bias.

5. Conclusion

In summary, we have demonstrated that management and I/O psychology meta-analytic reviews have largely ignored the issue of publication bias. After reviewing the causes of publication bias, we reviewed recommended methods for addressing it, including techniques never before used in our field (e.g., Egger's test of the intercept and cumulative meta-analysis). We then used these methods to assess publication bias in the meta-analytic review by Berry et al. (2010). Finally, we support the recommendations of others (e.g., American Psychological Association, 2010; Banks & McDaniel, 2011; Geyskens et al., 2009; McDaniel et al., 2006; Rothstein et al., 2005b) to include rigorous publication bias analyses in all meta-analytic reviews and to report the results regardless of the outcome. Likewise, we recommend that all researchers and test publishers follow the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) and *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 2003) and report all validity results available to them, regardless of the magnitude and direction of the correlations. The confidence we have in meta-analytic results is reliant upon the extent to which our research is free of publication bias.

Notes

1. Discussion of the issue of publication bias has existed in the social sciences for over three decades (e.g., Rosenthal, 1979) and more specifically, within the organizational sciences for 15 years prior to the start (i.e., 2005) of our review (e.g., Hunter & Schmidt, 1990, pp. 506–515). Advanced methods have also been in existence for over a decade in many cases (e.g., Begg & Mazumdar, 1994; Duval & Tweedie, 2000a, b; Egger et al., 1997).
2. We did not find any meta-analytic reviews published in *Administrative Science Quarterly* or *Organization Science*. We found one study in *Strategic Management Journal* that conducted a source comparison of published and unpublished studies.
3. We thank Christopher Berry for providing the data in an automated form. We note that the omission of the APA-recommended publication bias analyses in the Berry et al. article might be due to the article being accepted prior to the release of the 2010 APA style manual.

References

- References marked with an asterisk indicate studies included in the meta-analysis.
- Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011a). Meta-analytic choices and judgment calls: Implica-

- tions for theory building, testing, obtained effect sizes, and scholarly impact. *Journal of Management*, 37, 5–38.
- Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. (2011b). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods*, 14, 306–331.
- Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of organizational research methods: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, 12, 69–112.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Banks, G. C., Batchelor, J. H., & McDaniel, M. A. (2010). Smarter people are (a bit) more symmetrical: A meta-analysis of the relationship between intelligence and fluctuating asymmetry. *Intelligence*, 38, 393–401.
- Banks, G. C., & McDaniel, M. A. (2011). The kryptonite of evidence-based I-O psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 4, 40–44.
- Beal, D. J., Corey, D. M., & Dunlap, W. P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*, 87, 583–589.
- Becker, B. J. (2005). The failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111–126). Chichester, UK: Wiley.
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088–1101.
- Berlin, J. A., & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analysis. In H. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 35–48). Chichester, UK: Wiley.
- Berry, C. M., Sackett, P. R., & Tobares, V. (2010). A meta-analysis of conditional reasoning tests of aggression. *Personnel Psychology*, 63, 361–384.
- *Bing, M. N., Stewart, S. M., Davison, H. K., Green, P. D., McIntyre, M. D., & James, L. R. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology*, 92, 722–744.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- Briner, R. B., & Rousseau, D. M. (2011). Evidence-based I-O psychology: Not there yet. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 4, 3–22.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Davis, M. S. (1971). That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, 1, 309–344.
- Dickersin, K. (2005). Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein & A. J. Sutton (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 11–34). Chichester, UK: Wiley.
- Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 125–144). Chichester, UK: Wiley.
- Duval, S. J., & Tweedie, R. L. (2000a). A non-parametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89–98.
- Duval, S. J., & Tweedie, R. L. (2000b). Trim and fill: A simple funnel plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 276–284.
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634.
- Evangelou, E., Trikalinos, T. A., & Ioannidis, J. P. (2005). Unavailability of online supplementary scientific information from articles published in major journals. *FASEB Journal*, 19, 1943–1944.
- Evans, S. (1996). Statistician's comments (to Misleading meta-analysis: 'Fail safe N' is a useful mathematical measure of the stability of the results by R. Persaud). *British Medical Journal*, 312, 125.
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17, 120–128. doi: 10.1037/a0024445.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161–180.
- Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods*, 10, 444–467.
- *Frost, B. C. (2002). *Implicit versus explicit aggression: Who's ready to rumble?* Unpublished manuscript, Knoxville, TN: University of Tennessee.
- *Frost, B. C. (2005). *Implicit and explicit personality: An integrative approach to predicting aggressive behavior in a field setting*. Doctoral dissertation, University of Tennessee. ProQuest Dissertations, AAT 3188875.
- Geyskens, I., Krishnan, R., Steenkamp, J., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management*, 35, 393–419.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1–20.
- Hafdahl, A. R. (2009). Improved Fisher z estimators for univariate random-effects meta-analysis of correlations. *British Journal of Mathematical and Statistical Psychology*, 62, 233–261.
- Hafdahl, A. R. (2010). Random-effects meta-analysis of correlations: Monte Carlo evaluation of mean estimators. *British Journal of Mathematical and Statistical Psychology*, 63, 227–254.
- Hafdahl, A. R., & Williams, M. A. (2009). Meta-analysis of correlations revisited: Attempted replication and extension of Field's (2001) simulation studies. *Psychological Methods*, 14, 24–42.
- *Hawes, S. R. (2000). A comparison of biodata, ability, and a conditional reasoning test as predictors of reliable behavior

- in the workplace. Doctoral dissertation, University of Tennessee. ProQuest Dissertations, AAT 9996356.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Higgins, J. P., & Green, S. (2009). *Cochrane handbook for systematic reviews of interventions*. Version 5.0.2 [updated September 2009]: The Cochrane Collaboration. Available at <http://www.cochrane-handbook.org>.
- Hopewell, S., Clarke, M., & Mallett, S. (2005). Grey literature and systematic reviews. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 48–72). Chichester, UK: Wiley.
- *Hornick, C. W., Fox, K. A., Axton, T. R., & Wyatt, B. S. (1999). The relative contribution of conditional reasoning and multiple intelligence measures in predicting firefighter and law enforcement officer job performance. Unpublished manuscript.
- Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, 80, 327–334.
- Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, 75, 334–349.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Ioannidis, J. P. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Journal of the American Medical Association*, 279, 281–286.
- Ioannidis, J. P. (2005). Differentiating biases from genuine heterogeneity: Distinguishing artifactual from substantive effects. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 287–302). Chichester, UK: Wiley.
- Ioannidis, J. P. (2006). Evolution and translation of research findings: From bench to where? *PLoS Clinical Trials*, 2, 1–5.
- James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, 1, 131–163.
- James, L. R., & Mazerolle, M. D. (2002). *Personality at work*. Thousand Oaks, CA: Sage.
- *James, L. R., & McIntyre, M. D. (2000). *Conditional reasoning test of aggression test manual*. Knoxville, TN: Innovative Assessment Technology, LLC.
- James, L. R., McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M. et al. (2005). A conditional reasoning measure for aggression. *Organizational Research Methods*, 8, 69–99.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24, 602–611.
- Journal Citation Reports. (2010). Published by Thompson Reuters. Accessed via the ISI Web of Knowledge database.
- Lau, J., & Antman, E. M. (1992). Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine*, 327, 248–254.
- Lau, J., Schmid, C. H., & Chalmers, T. C. (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology*, 48, 45–57.
- *LeBreton, J. M. (2002). Use of differential framing to measure implicit social cognitions associated with aggression. Doctoral dissertation, University of Tennessee. ProQuest Dissertations, AAT 3054126.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up*. Boston, MA: Harvard University Press.
- McDaniel, M. A. (2009, April). Cumulative meta-analysis as a publication bias method. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, 59, 927–953.
- McKay, P., & McDaniel, M. A. (2006). A re-examination of Black–White mean differences in work performance: More data, more moderators. *Journal of Applied Psychology*, 91, 531–554.
- Moreno, S. G., Sutton, A. J., Turner, E. H., Abrams, K. R., Cooper, N. J., Palmer, T. M. et al. (2009). Novel methods to deal with publication biases: Secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. *British Medical Journal*, 339, b2981.
- Orwin, R. F. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8, 157–159.
- *Patton, T. W. (1998). Measuring personal reliability via conditional reasoning: Identifying people who will work reliably. Doctoral dissertation, University of Tennessee. ProQuest Dissertations, AAT 9936290.
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Bachrach, D. G. (2008). Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *Journal of Management*, 34, 641–720.
- Rosenthal, R. (1979). The ‘file drawer problem’ and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In H. Cooper, L. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 103–126). New York, NY: Russell Sage Foundation.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005a). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005b). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 1–7). Chichester, UK: Wiley.
- *Russell, S. M., James, L. R., & McIntyre, M. D. (2004, April). Predicting lying, cheating, and defiance in an internet-based environment. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago, IL.
- *Sablinski, C. J., & Mitchell, T. R. (2006, May). Post hoc analyses of interactions between the aggression questionnaire and conditional reasoning test of aggression. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Sackett, P. R., & Larson, J. R. (1990). Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 419–489). Palo Alto, CA: Consulting Psychologists Press.

- Schmidt, F. L., & Le, H. (2005). *Hunter & Schmidt's meta-analysis programs (version 1.1)*. Iowa City, IA: The University of Iowa.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: The Society for Industrial and Organizational Psychology.
- Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J. et al. (2010). Dissemination and publication of research findings: An updated review of related biases. *Health Technology Assessment*, 14, 1–220.
- Stern, J. M., & Simes, R. J. (1997). Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal*, 315, 640–645.
- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). Chichester, UK: Wiley.
- Sutton, A. J. (2005). Evidence concerning the consequences of publication and related biases. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 175–192). Chichester, UK: Wiley.
- Sutton, A. J. (2009). Publication bias. In H. Cooper, L. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435–472). New York, NY: Russell Sage Foundation.
- Tate, B. W., & McDaniel, M. A. (2008). *Race differences in personality: An evaluation of moderators and publication bias*. Paper presented at the Annual meeting of the Academy of Management, Anaheim CA.
- Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, 22, 2113–2126.
- Trikalinos, T. A., & Ioannidis, J. P. A. (2005). Assessing the evolution of effect sizes over time. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 241–259). Chichester, UK: Wiley.
- Uitterlinden, A. G., Ralston, S. H., Brandi, M. L., Carey, A. H., Grinberg, D., Langdahl, B. L. et al. (2006). The association between common vitamin d receptor gene variations and osteoporosis: A participant-level meta-analysis. *Annals of Internal Medicine*, 145, 255–264.
- *Walton, W. R. (2004). Justification of antisocial behavior. Doctoral dissertation, University of Tennessee. ProQuest Dissertations, AAT 3148313.