
Debunking Myths and Urban Legends About Meta-Analysis

Organizational Research Methods
14(2) 306-331
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1094428110375720
<http://orm.sagepub.com>



Herman Aguinis¹, Charles A. Pierce²,
Frank A. Bosco², Dan R. Dalton¹, and
Catherine M. Dalton¹

Abstract

Meta-analysis is the dominant approach to research synthesis in the organizational sciences. We discuss seven meta-analytic practices, misconceptions, claims, and assumptions that have reached the status of myths and urban legends (MULs). These seven MULs include issues related to data collection (e.g., consequences of choices made in the process of gathering primary-level studies to be included in a meta-analysis), data analysis (e.g., effects of meta-analytic choices and technical refinements on substantive conclusions and recommendations for practice), and the interpretation of results (e.g., meta-analytic inferences about causal relationships). We provide a critical analysis of each of these seven MULs, including a discussion of why each merits being classified as an MUL, their kernels of truth value, and what part of each MUL represents misunderstanding. As a consequence of discussing each of these seven MULs, we offer best-practice recommendations regarding how to conduct meta-analytic reviews.

Keywords

meta-analysis, philosophy of science, quantitative research, literature review

Meta-analysis is the methodology of choice to synthesize existing empirical evidence and draw science-based recommendations for practice in the organizational sciences and many other fields. As such, meta-analysis has also been described as the “critical first step in the effective use of scientific evidence” (Rousseau, Manning, & Denyer, 2008, p. 476). An important reason for the prominence of meta-analysis is summarized in the iconic observation by Hunter, Schmidt, and Jackson (1982) 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).

The reliance on and growth of meta-analytic applications for conducting quantitative literature reviews have been extraordinary. Consider, for example, the period from 1970 through 1985. Over

¹ Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, Bloomington, USA

² Department of Management, Fogelman College of Business & Economics, University of Memphis, TN, USA

Corresponding Author:

Herman Aguinis, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Suite 630D, Bloomington, IN 47405, USA
Email: haguinis@indiana.edu

that 15-year period, the PsycINFO database includes 224 articles with the expression “meta-analysis” or its derivatives in the title, the Academic/Business Source Premier (EBSCO) database has 55 such articles, and there are also 55 such articles included in MEDLINE with the expression “meta-analysis” or its derivatives in the title. Compare those numbers with the more recent 15-year period (1994–2009): There are 3,481 articles in PsycINFO, 6,918 articles in EBSCO, and 11,373 articles in MEDLINE. There is additional evidence that the popularity of meta-analysis in the organizational sciences continues to accelerate at a rapid pace. Consider a recent review of meta-analyses published in *Academy of Management Journal* (AMJ), *Journal of Applied Psychology* (JAP), *Journal of Management* (JOM), *Personnel Psychology* (PPsych), and *Strategic Management Journal* (SMJ) from 1982 through August 2009 (Aguinis, Dalton, Bosco, Pierce, & Dalton, in press). Extrapolating into the future based on the empirically derived trend using data from 1982 through 2009 led to the prediction of Aguinis, Dalton, et al. (in press) that the number of meta-analytically derived effect sizes to be reported annually in just these five journals will approach 1,000 around the year 2015 and 1,200 around the year 2020.

The present article is about meta-analytic practices, misconceptions, claims, assumptions, and “things we just know to be true” (Lance, Butts, & Michels, 2006, p. 202) that have reached the status of myths and urban legends (MULs). It has been suggested that there are many such MULs in organizational research, several of which have been passed on as “received doctrines” (Lance & Vandenberg, 2009, p. 1). It is in that spirit that we provide a critical examination of seven MULs specifically in the domain of meta-analysis together with a discussion of their kernels of truth value (Vandenberg, 2006) as well as what part of them represents misunderstandings. To do so, we greatly benefited, and borrowed liberally, from the compendia and reviews of meta-analysis that preceded us (e.g., Aguinis, Dalton et al., in press; Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, Hedges, & Valentine, 2009; Geyskens, Krishnan, Steenkamp, & Cunha, 2009; Hartung, Knapp, & Sinha, 2008; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Schmidt, Hunter, Pearlman, & Hirsh, 1985; Schulze, 2004). Similarly, our task was facilitated by a formidable body of work generated from analytic and simulation research and reviews of and commentaries on this research within and outside of the organizational sciences (e.g., Bobko & Roth, 2003; Cooper, 2010; Cooper & Hedges, 2009; Dalton & Dalton, 2005, 2008; Dieckmann, Malle, & Bodner, 2009; Egger, Smith, & Altman, 2001; Field, 2001; Glass, 1977; Higgins & Green, 2009; Petticrew, 2001; Rosenthal, 1991; Schmidt, 2008; Stangl & Berry, 2000; Sutton, Abrams, Ades, Copper, & Welton, 2009; Wanous, Sullivan, & Malinak, 1989; Weed, 2000).

The attention to meta-analysis provides an imposing literature. We recognize the daunting scope of this initiative. Accordingly, although the research and commentary on which we rely is representative of that body of work, it is by no means exhaustive. On every point, we could have cited more broadly and credited more outstanding relevant work. We regret that all of the work contributing to the ubiquity, influence, and promise of meta-analysis is not directly represented in our article. Similarly, our article focuses specifically on seven MULs. These MULs are diverse in terms of the topics they address as well as their longevity: Some (e.g., MUL #1: A Single Effect Size Can Summarize a Literature) were established during the early days of meta-analysis, whereas others (e.g., MUL #7: Meta-analytic Technical Refinements Lead to Important Scientific and Practical Advancements) have become prominent in recent years. Readers may be able to identify additional MULs about meta-analysis, and we hope that future research will address them.

Next, we present these seven MULs about meta-analysis, provide justification for their classification as an MUL, and critically analyze the kernels of truth value as well as any misunderstandings for each. As a consequence of this critical analysis and separation of truth value from myth, we also distill best-practice recommendations for meta-analytic practice. A summary of our analysis of each MUL is included in Table 1.

Table 1. Summary of Meta-Analysis Myths and Urban Legends: Kernels of Truth Value, Misunderstandings, and Recommendations for Meta-Analytic Practice

Myth and Urban Legend (MUL)	Kernels of Truth Value	Misunderstandings	Recommendations for Meta-Analytic Practice
MUL #1: A single effect size can summarize a literature	A point estimate, usually the mean effect size, as any summary statistic, provides an estimate of the overall direction and strength of a relationship across the studies included in the meta-analysis.	An examination of a single summary effect size to the exclusion of an examination of the variance of study-level effect sizes provides an incomplete picture because it fails to recognize the conditions under which a particular relationship may change in direction and strength.	Report summary effect sizes but also the variance around the overall estimate as well as moderator variables that may explain this across-study variance. For example, report the statistic I^2 , which is an estimate of the population parameter τ^2 (i.e., the variance of the true effect sizes).
MUL #2: Meta-analysis can make lemonade out of lemons: Meta-analysis allows researchers to gather a group of inconclusive and perhaps poorly designed studies and draw impressive conclusions with confidence	Holding primary-level study quality constant, meta-analysis allows researchers to draw more accurate conclusions than primary-level researchers because of larger samples and, hence, improved external validity and stability in the resulting effect-size estimates.	Meta-analyses are not immune to low-quality studies and resulting estimates will be biased if the quality of the primary-level studies included in the meta-analysis is erroneously assumed to be uniformly high.	Assess each primary-level study, and the resulting meta-analytically derived estimates, in terms of risk factors that can lead to biased results. If studies are excluded, be clear about the criteria for exclusion and these criteria should be established prior to data collection, logically consistent in terms of the goals of the study, as well as be taken into account when discussing the generalizability and/or limitations of meta-analytic results.
MUL #3: File drawer analysis is a valid indicator of possible publication bias	If the resulting failsafe N is small, there is a possibility that publication bias exists.	A large—however defined—failsafe N does not necessarily imply lack of publication bias.	Use the trim-and-fill method to assess potential publication bias.
MUL #4: Meta-analysis provides evidence about causal relationships	Meta-analysis can provide evidence regarding the plausibility of causal relationships by using meta-analytically derived correlation matrices as input for testing models positing competing causal relationships and assessing the consistency and temporality of a relationship across settings.	Phrases such as the “effect” or “impact” of a variable on another are subtle and implicit statements about causality, but claims about knowledge of causal relationships based on meta-analytic results are typically not justified.	Use meta-analysis to gather preliminary evidence, as well as produce hypotheses, regarding the possible causal relationship between variables.

(continued)

Table 1 (continued)

Myth and Urban Legend (MUL)	Kernels of Truth Value	Misunderstandings	Recommendations for Meta-Analytic Practice
MUL #5: Meta-analysis has sufficient statistical power to detect moderating effects	Because of its large sample size, a meta-analysis is likely to have greater statistical power than a primary-level study examining a similar research question.	Many factors in addition to small sample size have a detrimental effect on the statistical power of tests for moderating effects and, hence, such tests are usually performed at insufficient levels of statistical power.	Follow recommendations offered by Aguinis, Gottfredson, and Wright (in press) and perform a priori power calculations as described by Hedges and Pigott (2004).
MUL #6: A discrepancy between results of a meta-analysis and randomized controlled trials (RCTs) means that the meta-analysis is defective	In some cases it may be that a discrepancy can be due to problems in the design and/or execution of a meta-analysis such as the inclusion of poor-quality studies (see MUL #2).	In most cases, a discrepancy between results of a meta-analysis and some RCTs is to be expected because the meta-analytic summary estimate is an average of the primary-level effect sizes.	Do not focus on possible discrepancies between results of meta-analyses and RCTs; rather, use meta-analysis to explain across-study variability caused by artifactual sources of variance (e.g., sampling error) and substantive moderator variables.
MUL #7: Meta-analytic technical refinements lead to important scientific and practical advancements	Technical refinements lead to improved accuracy and estimation that may be meaningful in certain contexts only.	In most organizational science contexts, technical refinements do not have a meaningful and substantive impact on theory or practice.	Use the best and most accurate estimation procedures available, even if the resulting estimates are only marginally superior.

MUL #1: A Single Effect Size Can Summarize a Literature

The dual goals of a meta-analysis are to (a) estimate the overall strength and direction of an effect or relationship and (b) estimate the across-study variance in the distribution of effect-size estimates and the factors (i.e., moderator variables) that explain this variance (Borenstein et al., 2009; Hedges & Olkin, 1985). This second goal is, essentially, to investigate the possible presence of moderating effects: Conditions under which the size and even direction of an effect changes (Aguinis & Pierce, 1998; Aguinis, Sturman, & Pierce, 2008; Cortina, 2003; Sagie & Koslowsky, 1993). As noted by Borenstein et al. (2009), “the goal of a meta-analysis should be to synthesize the effect sizes, and not simply (or necessarily) to report a summary effect. If the effects are consistent, then the analysis shows that the effect is robust across the range of the included studies. If there is modest dispersion, then this dispersion should serve to place the mean effect in context. If there is substantial dispersion, then the focus should shift from the summary effect to the dispersion itself. Researchers who report a summary effect and ignore heterogeneity are indeed missing the point of the synthesis” (p. 378).

In spite of the dual goals of meta-analysis, there is an overreliance on a single effect size to summarize the primary-level studies included in a review, usually a mean effect size. In fixed-effect models, this mean effect size represents the mean of the true population effect sizes, whereas in random-effect models this mean represents the mean of the distribution of true population effect sizes. About 20% of meta-analyses published in 14 organizational science journals between 1980 and 2007 reported a mean effect size but did not report any analyses regarding potential heterogeneity of effect sizes across studies (Geyskens et al., 2009). This overreliance on a single summary effect size to the exclusion of a discussion of effect size variance across studies is even more pronounced in the case of authors referring to a previously published meta-analysis. Specifically, Carlson and Ji (2009) identified 253 citations to meta-analyses and examined how authors described and used the results of prior meta-analyses. The review by Carlson and Ji revealed the disturbing fact that the variability of effect sizes was not reported in any of the 253 journal articles citing previously published meta-analyses. Instead, it seems that it is more convenient to report a point estimate of the relationship between two variables. This type of reporting perhaps makes for a simpler and cleaner story, particularly when an author is trying to make a point regarding the nature of a particular relationship without going into a discussion of the conditions under which that relationship may change in strength or form. For example, Shedler (2010) reviewed previously published meta-analyses that examined the efficacy of psychodynamic psychotherapy and concluded that effect sizes for psychodynamic therapy are as large as those reported for other therapies. He reached this conclusion by reporting 18 meta-analytically derived mean and median effect sizes, but the variance was not reported for any of these 18 summary effects. Focusing on a single summary effect size to the exclusion of moderating effects (i.e., conditions under which the relationship changes) and possible mediating effects (i.e., processes underlying the direct relationship) is a hindrance to theory development (Aguinis, 2004; Hall & Rosenthal, 1991).

Studies included in a meta-analysis are rarely homogenous (Rousseau et al., 2008). This heterogeneity in the primary-level studies included in a meta-analysis has been an early target of criticism. For example, Eysenck (1984) noted that “adding apples and oranges may be a pastime for children learning to count, but unless we are willing to disregard the differences between these two kinds of fruit, the result will be meaningless” (p. 57). In colorful rebuttals, several authors have noted that heterogeneity in the pool of primary-level studies included in a meta-analysis is not a burden for the meta-analyst but, rather, it is an opportunity. Indeed, it has been noted that “variety is the spice of life” (Cooper & Hedges, 2009, p. 563) and that such diversity is “not only inevitable, but also desirable” (Borenstein et al., 2009, p. 358). The key issue is not whether a meta-analytic data set is heterogeneous—this is pretty much a fact of life in conducting a quantitative literature review. The key issue is to establish study inclusion criteria prior to data collection based on the substantive

goals of the meta-analysis. If the research question posed is broad, the inclusion criteria will also be broad and the resulting set of studies are likely to be quite heterogeneous. In short, Glass (1978) noted “of course they are comparing apples and oranges; such comparisons are the only endeavor befitting competent scholarship One compares apples and oranges in the study of fruit” (p. 395). Similarly, Rosenthal and DiMatteo (2001) wrote that “it is a good thing to mix apples and oranges, particularly if one wants to generalize about fruit” (p. 68).

Also related to setting inclusion criteria and the resulting heterogeneity of the primary-level studies in a meta-analysis, it is often the case that what seems to be the same research question can be posed at different levels of specificity and the level of specificity will determine choices regarding the inclusion of studies. For example, the question “Is employee withdrawal behavior related to job satisfaction?” would justify inclusion of a broader set of studies compared to the research question “Is employee turnover related to extrinsic satisfaction?” Accordingly, the degree of heterogeneity in a meta-analytic data set will be dictated in part by the types of research questions posed prior to data collection.

In sum, the MUL is that a single effect size can summarize a literature. The kernel of truth value is that a point estimate, usually a mean effect size, as any summary statistic, simply provides an estimate of the overall direction and strength of a relationship across the studies included in the meta-analysis. However, an examination of a single summary effect size to the exclusion of an examination of the variance of effect sizes provides an incomplete picture because it fails to recognize the conditions under which a particular relationship may change in direction and/or strength. Thus, in terms of meta-analytic practice, meta-analysts, as well as authors citing and referring to results of previously conducted meta-analyses, should report both summary effect sizes and the variance around the overall estimate as well as moderator variables that may explain this variance. For example, meta-analysts can compute the value for the statistic T^2 , which is an estimate of the population parameter τ^2 (i.e., the variance of the true effect sizes). Thus, T serves as an estimate of the standard deviation of true effects and can be used to create a 95% confidence interval around the summary effect size.

MUL #2: Meta-Analysis Can Make Lemonade out of Lemons

Hans Eysenck (1978), an early critic of meta-analytic reviews, was concerned that the input for a meta-analysis is often “a mass of reports—good, bad, and indifferent” (p. 517). He also reminded us that “garbage in, garbage out is a well-known axiom of computer specialists; it applies here [for meta-analysis] with equal force” (p. 517). Eysenck was reacting to the MUL, established early in the history of meta-analysis, that meta-analysis allows researchers to make lemonade out of lemons. In other words, the MUL is that meta-analysis allows researchers to gather a group of inconclusive and perhaps poorly designed and executed studies including, for example, small samples and unreliable measures, and yet draw impressive conclusions with confidence (Hunter et al., 1982; Schmidt, 1992).

The abiding issue, now as then, is the quality of the primary studies included in a meta-analytic database. Although this issue is not addressed regularly or systematically in the organizational sciences, meta-analytic reviews published in fields such as medicine, biology, nursing, and criminal justice, among others, include quality screens to make decisions about whether a particular study should be included in the meta-analytic database (for recent illustrations, see Almekhlafi et al., 2009; Hanson, Bourgon, Helmus, & Hodgson, 2009; Härlein, Dassen, Halfens, & Heinze, 2009; Kastner & Straus, 2008). The reason for establishing inclusion criteria based on quality—however defined—is that poor-quality studies are likely to affect the integrity of meta-analytically derived estimates (e.g., Cooper, 2010). Although the majority of studies included in a meta-analysis are usually drawn from peer-reviewed journals and, hence, have survived the review process and are the

product of a series of revisions, the peer review system is far from perfect (Bedeian, Van Fleet, & Hyman, 2009a, 2009b; Hitt, 2009; Klimoski, 2009; Tsui & Hollenbeck, 2009). Accordingly, it is likely that some low-quality and low-impact manuscripts are published even in what are considered to be the most prestigious journals (Starbuck, 2005).

The concern regarding the inclusion of poor-quality studies in a meta-analysis suggests the necessity of a methodological practice with the goal of deciding whether a particular study should be included and, if included, the particular weight it should be assigned. The challenge with this type of meta-analytic practice is that many scales used to rate a study's quality are not highly correlated with each other and this low correlation is obtained even when ratings are provided by trained evaluators (Jüni, Witschi, Bloch, & Egger, 1999). For example, Gottfredson (1978) found that interrater agreement on study quality for expert evaluators was in the .40s. This finding is consistent with the general poor degree of agreement among reviewers involved in the peer review process, which is approximately .30 or .40 (Gilliland & Cortina, 1997). Moreover, there is the concern that even when quality scales are used systematically and reliably, their construct validity may still be suspect (Valentine, 2009; Valentine & Cooper, 2008).

Fortunately, there is an alternative for addressing primary-level study quality issues that does not include the use of quality scales that are suspect both on reliability and validity grounds. This alternative is a practice that has been adopted in health care and other fields outside of the organizational sciences, and it involves an assessment of specific features of primary-level studies that affect the risk of obtaining biased results. Many of these risk factors are based on the familiar threats to validity described by Shadish, Cook, and Campbell (2002).

One useful list of risk factors that can potentially lead to biased results has been made available by the Cochrane Collaboration (see Higgins & Green, 2009), which is "an international not-for-profit and independent organization, dedicated to making up-to-date, accurate information about the effects of healthcare readily available worldwide" (<http://www.cochrane.org/docs/descrip.htm>). Consider the following examples of criteria used regarding the potential risk factor "allocation concealment." This risk factor involves the possibility that meta-analytic results may be biased because researchers who conducted the primary-level studies unconsciously or otherwise influenced which participants were assigned to a given intervention group (the complete list of risk factors and criteria is available online at <http://www.ohg.cochrane.org/forms/Risk%20of%20bias%20assessment%20tool.pdf>):

Low risk of bias:

Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:

- central allocation (including telephone, web-based, and pharmacy-controlled, randomization)
- sequentially numbered drug containers of identical appearance
- sequentially numbered, opaque, sealed envelopes

High risk of bias:

Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on the following:

- using an open random allocation schedule (e.g., a list of random numbers)
- using assignment envelopes without appropriate safeguards (e.g., if envelopes were unsealed or nonopaque or not sequentially numbered)

- alternation or rotation
- date of birth
- case record number
- any other explicitly unconcealed procedure

This example refers to a general intervention that could be, for example, the deployment of a new training program (Aguinis, Mazurkiewicz, & Heggstad, 2009). In this particular training intervention, employees were randomly assigned to the training (i.e., new web-based procedure) and control groups. This primary-level study would be classified as having low risk of bias because assignment to one or the other group was done from a central location and employees were numbered sequentially and assigned to one or the other group based on having an odd or even number. Regarding potential high risk of bias, investigators did not assign numbers to individuals (i.e., this process was done automatically online), and no information regarding date of birth or any other employee identifier was recorded.

Finally, we emphasize that criteria for exclusion based on methodological limitations should be established prior to data collection rather than during the process of collecting studies to be included in a meta-analysis. Setting criteria a priori minimizes the possibility of introducing publication bias by eliminating studies with which a meta-analyst does not agree with the conscious or unconscious pretext that they are “methodologically weak.” Of course, sometimes a specific study will require revisiting the prespecified criteria but, for the most part, the criteria should be set at the predata collection stage.

In sum, the MUL is that meta-analysts can make lemonade out of lemons: There is the belief that meta-analysis allows researchers to gather a group of inconclusive and perhaps poorly designed studies and draw impressive conclusions with confidence. The kernel of truth value is that, holding primary-level study quality constant, meta-analysis allows researchers to draw more accurate conclusions than primary-level researchers because of larger samples and, hence, improved external validity and stability in the resulting estimates. However, meta-analyses are not immune to low-quality studies and resulting estimates will be biased if the quality of the primary-level studies included in the meta-analysis is erroneously assumed to be uniformly high. Thus, meta-analysts need to assess each primary-level study, and the resulting meta-analytically derived estimates, in terms of risk factors that can lead to biased results. Essentially, these risk factors are methodological features of the primary-level studies related to the familiar threats to validity (Shadish et al., 2002). In addition, if studies are excluded, meta-analysts need to be clear about the criteria for exclusion and these criteria should be established prior to data collection, logically consistent in terms of the goals of the study, as well as be taken into account when discussing the generalizability and/or limitations of meta-analytic results.

MUL #3: File Drawer Analysis is a Valid Indicator of Possible Publication Bias

Publication bias occurs when “the research that appears in the published literature is systematically unrepresentative of the population of completed studies” (Rothstein, Sutton, & Borenstein, 2005, p. 1). Most meta-analyses are comprised largely of studies that are published in the scientific literature (i.e., peer-reviewed journals) because unpublished studies are often hiding in researchers’ file drawers and electronic media storage. An important difference between published and unpublished studies is that, compared to published ones, unpublished studies may be less likely to report successful and/or statistically significant results because, as noted in an editorial published in August 1909 in *The Boston Medical and Surgical Journal*, “it is natural that men [sic] should be eager to present to the world their successes rather than their failures” (Dickersin, 2005, pp. 11–12). Thus, an

important type of publication bias occurs when a meta-analysis excludes relevant studies, possibly unpublished, which report results that are not statistically significant or seen as unsuccessful (i.e., small in magnitude) from an intervention or therapeutic standpoint (Begg, 1994) or that contravene financial, political, ideological, professional, or other interests of investigators, research sponsors, and journal editors (Halpern & Berlin, 2005).

Although the availability of Internet LISTSERVS for posting e-mail queries about locating unpublished studies has helped to address this challenge, publication bias is still an important threat to the validity of meta-analytic results (Dickersin, 2005). Specific examples of publication bias include the tendency not to publish studies reporting null results and the tendency to publish studies reporting small rather than large race-based differences (Borenstein et al., 2009; McDaniel, Rothstein, & Whetzel, 2006; Sutton, 2009). Moreover, the threat of selective outcome reporting may be greater than the threat of studies that are completely missing. This is because in the organizational sciences researchers rarely test a single hypothesis or bivariate relationship (as they do in intervention studies in the medical and biological sciences). Rather, each study includes tests of several relationships. It is unlikely that none of the hypotheses will receive support but quite likely that a subset will not. Thus, the tendency may be for authors to selectively report those relationships where a hypothesis is confirmed and to omit information about outcomes or relationships that were not confirmed (Chan, Hróbjartsson, Haahr, Gøtzsche, & Altman, 2004; McDaniel et al., 2006; Pigott, 2009; Russell et al., 1994).

To address the concern that publication bias may affect meta-analytic results, Rosenthal (1979) proposed a solution to what he called the “file drawer problem.” Rosenthal’s solution entails conducting a file drawer analysis whereby the meta-analyst calculates the number of unpublished studies with statistically nonsignificant effects that would be needed to nullify (i.e., render statistically nonsignificant) the meta-analytically derived mean effect size. The number of required unpublished studies reporting a null effect is called the failsafe N . A small—however defined—value for a failsafe N suggests that the meta-analytic results may be affected by publication bias (Becker, 2005).

Becker (2005) noted that the failsafe N “was one of the earliest approaches for dealing with the problem of publication bias and, in the social sciences, is still one of the most popular” (p. 111). Moreover, numerous organizational science meta-analysts have assumed that using a failsafe N to determine whether their results were affected by publication bias is sufficient to minimize publication bias. If the obtained failsafe N is large—however defined—meta-analysts draw the inference that results are valid, replicable, difficult to disprove, and even practically important and meaningful. Consider the following verbatim statements from meta-analyses published in *Journal of Applied Psychology* and *Personnel Psychology* as recently as 2007 (because statements such as the ones below are common and we do not wish to single out any particular authors, citations for each of these sources are available on request):

- “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 $\pm .10$ (i.e., a small correlation with lower practical significance, per Cohen, 1969). These results appear in Table 4, demonstrating that the current findings are unlikely to be significantly affected by publication bias.”
- “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.”
- “In the present review, 3,450 unpublished studies with disconfirming results would be needed to invalidate the conclusions.”

In recent years, however, researchers have expressed their frustration that “publication bias is easier to detect than to correct” (Cooper & Hedges, 2009, p. 565; see also Rothstein et al., 2005; Sutton,

2005). More importantly, there are limitations of Rosenthal's file drawer method using a failsafe N , and hence it is no longer a recommended method for assessing publication bias in meta-analysis (McDaniel et al., 2006; Sutton, 2009). Among these limitations are (a) the assumption that excluded studies show a null result whereas many may, instead, show a result in the opposite direction, (b) ignoring primary-level study sample size information (e.g., assuming that the effect of adding N studies showing a null effect would be the same if each of the primary-level studies had a sample size of 10 or 10,000), and (c) the lack of a definition regarding what is a tolerable failsafe N value (Becker, 2005).

In 2006, McDaniel et al. introduced researchers in the organizational sciences to the trim-and-fill method of publication bias analysis. This methodology has been described in the health care literature by Duval and Tweedie (2000a, 2000b) and has since been shown to outperform a file drawer analysis for the purpose of assessing publication bias (see McDaniel et al., 2006 for a discussion of advantages of the trim-and-fill method over the file drawer method). The trim-and-fill method is a type of sensitivity analysis because it estimates how a summary effect size would change by adding potentially missing studies (Duval, 2005) and involves the following three steps:

- (1) By inspecting a funnel plot of study-level effect-size estimates on the horizontal axis against a measure of study size on the vertical axis, a meta-analyst can determine whether there is (when the plot resembles an asymmetrical inverted funnel) or is not (when the plot resembles a symmetrical inverted funnel) possible publication bias (Sterne, Becker, & Egger, 2005). In addition to a visual examination of a funnel plot, formal statistical tests of asymmetry include a rank correlation test by Begg and Mazumdar (1994) and a linear regression test by Egger, Davey Smith, Schneider, and Minder (1997).
- (2) If the funnel plot is asymmetrical, and thus suggests potential publication bias, then the trim-and-fill method imputes the missing study-level effect-size estimates needed to make the funnel plot symmetrical, adds them to the meta-analysis, and calculates a trim-and-fill adjusted mean effect-size estimate.
- (3) To assess the potential impact of publication bias, the meta-analyst examines the size of the difference between the value of the observed mean effect-size estimate and the value of the trim-and-fill adjusted mean effect-size estimate. Thus, the main goal of the trim-and-fill method is not to find the values of missing studies but, rather, to assess how much the value of the estimated summary effect size might change if there are missing studies (Duval, 2005).

Considering that the trim-and-fill method has only recently become the recommended technique for publication bias analysis in the organizational sciences, it is important to note that publication bias is not the only source of asymmetry in funnel plots. As summarized by Egger et al. (1997), sources of asymmetry include publication bias but also true heterogeneity (i.e., substantive factors related to effect size), artifacts (e.g., heterogeneity due to poor choice of effect measure), data irregularities (e.g., poor methodological design of small studies, inadequate analysis, fraud), and chance. In addition, although the trim-and-fill method seems to be the best current technique for detecting potential publication bias, it does have some limitations and the most important one is that it can confuse true heterogeneity with bias (Terrin, Schmid, Lau, & Olkin, 2003). Therefore, although it is highly recommended for use with relatively homogeneous subgroups of effect sizes, it may lead to erroneous conclusions when used with an entire set of effect sizes or with the subgroups that include a small number of effect sizes only (Peters, Sutton, Jones, Abrams, & Rushton, 2007).

In sum, the MUL is that a file drawer analysis resulting in a large failsafe N value is an effective and sufficient methodological practice to resolve the problem of publication bias in meta-analysis. The kernel of truth value is that if the resulting failsafe N is low, there is a good possibility that publication bias exists. However, the misunderstanding is that a large failsafe N does not necessarily

imply lack of publication bias. In terms of best meta-analytic practices, the recommendation is to use the trim-and-fill method to assess the effect of possible publication bias.

MUL #4: Meta-Analysis Provides Evidence About Causal Relationships

There is a great divide separating science and practice in management and related fields (Aguinis, Werner, Abbott, Angert, Park, & Kohlhausen, 2010; Cascio & Aguinis, 2008; Rynes, Bartunek, & Daft, 2001). A favorable feature of meta-analysis is that it can play a key role in bridging this widely documented science–practice gap. Consider the role of meta-analysis in the evidence-based management movement, which is the systematic use of the best available evidence to improve management practice (Pfeffer & Sutton, 2006; Rousseau et al., 2008). Meta-analysis is central to evidence-based management because it represents “a key methodology for locating, appraising, synthesizing, and reporting best evidence” (Briner, Denyer, & Rousseau, 2009, p. 24). Moreover, meta-analysis is seen as a methodological tool that allows researchers to draw causal inferences from data. Obviously, being able to draw causal inferences is just as important in the organizational sciences as in any other field concerned with the relative value of interventions. For example, a meta-analysis conducted by Brind, Chinchilli, Severs, and Summy-Long (1996) concluded that several thousand breast cancer deaths each year can be attributed to induced abortions. Similarly, strong claims about causality were made by Sood (1991) based on a meta-analysis regarding the effect of cigarette smoking on cervical cancer.

Claims about causality also abound in meta-analyses published in the organizational sciences. Most of these claims about causality are subtle and refer to the “impact of” or “effect of” a variable on another. By choosing to use language referring to “effect” or “impact” instead of “relationship,” “association,” or “covariation,” meta-analysts imply, and subsequently readers infer, that there is a causal link. As a few illustrations, consider the following verbatim statements extracted from meta-analyses published in 2009 in the *Journal of Applied Psychology* (because statements such as the ones below are common and we do not wish to single out any particular authors, citations for each of these sources are available on request):

- “results suggest that high-fidelity mock interviews may in fact be a useful method for learning about the nature of the effect that self-presentation tactics have on the interviewer.”
- “diversity with regard to job-related attributes has a greater impact on performance than diversity with regard to less job-related attributes.”
- “Examining the impact of college interventions on psychosocial mediators, we found that overall interventions had moderate to strong effects on motivational control and emotional control.”
- “Although we found information sharing had the strongest impact on performance on intellectual hidden profile tasks, information sharing also positively affected performance on less demonstrable as well as nonhidden profile tasks.”

In spite of these claims, in most cases meta-analysis does not allow researchers to make inferences about causality because a meta-analysis is a passive observational study. Because the majority of studies in the organizational sciences use a cross-sectional design (Aguinis, Pierce, Bosco, & Muslin, 2009; Podsakoff & Dalton, 1987), meta-analysts also use databases that contain mostly cross-sectional designs. However, even if the primary-level studies implemented an experimental design, the resulting meta-analysis is a passive observational study (Borenstein et al., 2009, pp. 209–210). For example, a meta-analysis investigating whether two types of training programs have differential effects on learning may include experimental primary-level studies that examined the effects of training type A and other experimental primary-level studies that examined the effects of training type B. A meta-analytic subgroup analysis may suggest a difference between the two

types of training. However, this difference could be caused by the type of training but also a host of other factors such as different individual characteristics of the trainees across the subsets of primary-level studies. The only situation in which a meta-analysis can provide direct evidence regarding causality is the very rare scenario when all the primary-level studies (a) implemented an experimental design and (b) are identical in all respects (e.g., participants, research team conducting the experiments, experimental manipulation, data collection procedures, and so forth). We are not aware of any published meta-analysis in the organizational sciences that meets these conditions.

Although meta-analysis alone cannot provide definitive evidence regarding causal relationships, it can play an important role and make valuable contributions in terms of providing preliminary evidence and also hypothesizing such relationships, which can subsequently be tested using experimental research designs. Consider the following three contributions.

First, meta-analytically derived effect-size estimates can be used as input in subsequent path-analytic or structural equation model testing (Bergh, Aguinis, Hanke, & Perry, 2010; Colquitt, LePine, & Noe, 2000; Viswesvaran & Ones, 1995). In other words, a meta-analytically derived correlation matrix can be used as input to test the fit of competing models such that in one model variable A is hypothesized to be an antecedent to variable B and in a second model variable B is hypothesized to be an antecedent to variable A. Indices of fit as well as a chi-square statistic can be computed to understand whether the meta-analytically derived input matrix has a greater correspondence with the matrix implied by the first or second model (Aguinis & Harden, 2009). These results would not provide definitive evidence regarding causation, particularly when most or all studies included in a meta-analytic database used a cross-sectional design, but would suggest which of the two causal directions is more plausible. Moreover, this type of model testing has great potential in terms of producing important theoretical breakthroughs (Rodgers, 2010; Vandenberg & Grelle, 2009). Because definitive proof regarding causality is virtually impossible, the state-of-the-science methodological approach is to assess the relative fit of the proposed causal chain to the data (Rodgers, 2010; Vandenberg & Grelle, 2009). Thus, as noted by González-Benito, Aguinis, Boyd, and Suárez-González (2010), "evidence regarding causality is established in relationship to other hypothesized causal chains and not in absolute terms." We acknowledge that in spite of its potential, there are some unresolved issues regarding the use of path-analysis and structural equation modeling with a meta-analytically derived input matrix. These unresolved issues include the use of correlations instead of covariances, possible presence of empty cells, heterogeneity of effect sizes within each cell, and the determination of sample size to use in tests of significance of the paths in the model (Cheung, 2008; Furlow & Beretvas, 2005).

Second, meta-analysis is able to provide evidence regarding the consistency of a particular relationship across settings. Consistency is an important criterion in terms of establishing causal relationships. Specifically, a highly homogeneous relationship between two variables across different settings and samples suggests that these variables may be causally related to each other (Weed, 2000). However, lack of cross-study homogeneity does not mean lack of causation because this may simply suggest the presence of moderators (see MUL #1).

Third, meta-analysis is able to provide evidence regarding temporality in a specific relationship. For example, also related to the earlier point about model testing, a meta-analyst can subgroup experimental studies that examined the effect of A on B and experimental studies that examined the effect of B on A. A comparison of mean effect sizes, as well as homogeneity of effect sizes, for each of these subgroups would provide evidence about whether the effect of A on B is larger or smaller than the effect of B on A.

In sum, the MUL is that meta-analytic results provide evidence about causal relationships: Authors of meta-analyses published in the organizational sciences often refer to the "effect" or "impact" of one variable on another. These claims about causality are often not explicit and in many cases they are subtle and implicit but, in most cases, they are not justified. The kernel of truth value

is that meta-analysis can provide evidence regarding the plausibility of causal relationships. Specifically, meta-analytically derived correlation matrices can be used as input for testing models positing competing causal relationships, and meta-analysis can be used to assess the consistency and temporality of a relationship across settings. Thus, meta-analysis can serve a useful role in gathering preliminary evidence, as well as producing hypotheses, regarding the possible causal relationship between variables.

MUL #5: Meta-Analysis Has Sufficient Statistical Power to Detect Moderating Effects

Meta-analysis is not only concerned with the overall strength and direction of an effect or relationship but also with the across-study variance in the effect or relationship estimates and the factors (i.e., moderator variables) that explain such variance if it exists. Accordingly, the process for estimating possible moderating effects includes two steps. First, a meta-analyst assesses whether effect sizes vary across primary-level studies (i.e., degree of heterogeneity or dispersion of effect sizes). Second, if across-study heterogeneity is found, and it is not explained by artifactual sources such as sampling error, then follow-up analyses are conducted to assess the presence of moderator variables associated with this variance.

If there is heterogeneity and the moderator variable is categorical, the next step is to conduct a subgroup analysis in which each study is assigned a numerical value based on the moderator (e.g., gender: 1 = female, 2 = male) and grouped according to this coding scheme. Then, the within-group effect sizes are computed and the presence of a moderator is confirmed if there is between-group heterogeneity. Although there are several types of subgroup analysis, they are all algebraically equivalent and yield the same p values (Borenstein et al., 2009, Chapter 19). If there is heterogeneity and the moderator variable is continuous, then the next step, instead of subgroup analysis, is to conduct a meta-regression analysis. Similar to multiple regression, meta-regression consists of a regression model involving predictors (i.e., potential moderators) and a criterion (i.e., weighted effect sizes). Just like multiple regression, meta-regression can also accommodate categorical predictors. Assuming a situation with two continuous moderators, the meta-regression model is $\mathbf{ES} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \boldsymbol{\varepsilon}$, where \mathbf{ES} represents a vector of weighted effect size values, \mathbf{X}_1 represents a vector of values for the hypothesized moderator X_1 , \mathbf{X}_2 represents a vector of values for the hypothesized moderator X_2 , and $\boldsymbol{\varepsilon}$ represents a vector of residuals. A Z test is then conducted to test the null hypothesis that each of the regression coefficients is zero in the population. In addition, confidence intervals can be computed around each meta-regression coefficient (Bonett, 2008).

In general, researchers assume that meta-analysis has adequate statistical power to detect moderator variables (Muncer, Taylor, & Craigie, 2002). In other words, it is assumed that if heterogeneity of effect sizes exists in the population, such heterogeneity will not be underestimated in a meta-analytic database. Moreover, there is the belief that, in most cases, researchers are more likely to err in the direction of overestimating as opposed to underestimating across-study heterogeneity (Aguinis, 2001; Aguinis & Whitehead, 1997). As noted by Hunter and Schmidt (2004), “the corrected standard deviation of results across studies should always be regarded as an overestimate of the true standard deviation” (p. 66). In addition, it is generally assumed that if heterogeneity is observed, there will be sufficient statistical power to detect differences between subgroups (in the case of categorical moderators) and there will be sufficient statistical power to detect nonzero regression coefficients (in the case of continuous moderators).

Our position that these two assumptions are in fact MULs is based on the following evidence. First, regarding the overall across-study heterogeneity, the variance of the true population effect sizes is denoted by the parameter τ^2 and its sample-based estimate T^2 when using the meta-analytic procedures developed by Hedges and colleagues (Borenstein et al., 2009; Hedges

& Olkin, 1985; Hedges & Vevea, 1998; the Hunter & Schmidt, 2004, procedures include a similar estimate which is computed as the total observed variance minus the variance estimated to be caused by methodological and statistical artifacts). The size of T^2 depends not only on the true variance of effect sizes in the population but also on the size of the confidence interval for the effect sizes included in the meta-analysis. Thus, as noted by Borenstein et al. (2009), “if the studies themselves have poor precision (wide confidence intervals), this could mask the presence of real (possibly substantively important) heterogeneity, resulting in an estimate of zero for T^2 ” (p. 122). The fact that meta-analyses generally do not report confidence intervals for τ^2 suggests the implicit assumption that heterogeneity estimates are precise and accurate. Second, regarding follow-up moderator tests when heterogeneity is found, results of numerous Monte Carlo studies indicate that such tests usually have insufficient statistical power (Aguinis et al., 2008; Sackett, Harris, & Orr, 1986; Sagie & Koslowsky, 1993). However, meta-analyses generally do not report results of a priori statistical power analyses, even in cases when evidence is not found in support of moderators.

One possible reason why statistical power is assumed to be sufficient in testing moderating effect hypotheses meta-analytically is that sample sizes are substantially larger compared to primary-level studies. Given that sample size is an important determinant of statistical power (Cohen, 1988) and a larger sample size will reduce the standard error of the weighted mean effect size in fixed-effect models (Cohn & Becker, 2003), meta-analysts and meta-analysis consumers may simply assume that adequate power is guaranteed. However, small sample size is only one of the many factors that have a detrimental effect on statistical power for moderator variable tests (Aguinis, 2004; Aguinis, Beaty, Boik, & Pierce, 2005; Aguinis, Culpepper, & Pierce, 2010; Aguinis & Stone-Romero, 1997; Shieh, 2009). Moreover, another important factor that affects power in a meta-analysis is the number of primary-level studies included in the database. So, even if a meta-analysis includes a sample size in the thousands, a small number of primary-level studies may lead to insufficient statistical power to detect across-study heterogeneity. In addition, other factors that also have a detrimental impact on power include variable truncation, measurement error, scale coarseness, and unequal proportions across subgroups in the case of categorical moderators (Aguinis, Pierce, & Culpepper, 2009).

In sum, the MUL is that meta-analysis has sufficient statistical power to detect moderating effects. In other words, it is assumed that the true population effect size heterogeneity will be assessed accurately; also, it is assumed that if heterogeneity exists, it will be accurately associated with particular moderator variables. The kernel of truth value is that, because of its large sample size, a meta-analysis is likely to have greater statistical power than a primary-level study examining a similar research question. However, there are many factors in addition to sample size that have a detrimental effect on the statistical power for tests of hypothesized moderating effects. Meta-analysts need to be aware that power is likely to be insufficient in many situations. This is particularly detrimental for the advancement of the organizational sciences because once a conclusion of “no moderation” is reached meta-analytically, it is unlikely that follow-up primary level research will attempt to find such moderating effects and contradict a previously published meta-analysis. In terms of recommendations for practice regarding the estimation of moderating effects, meta-analysts can follow 13 recommendations offered by Aguinis, Gottfredson, and Wright (in press) regarding actions researchers can take before and after data collection and also perform a priori power calculations as described by Hedges and Pigott (2004).

MUL #6: A Discrepancy Between Results of a Meta-Analysis and Randomized Controlled Trials Indicates That the Meta-Analysis Is Defective

Randomized controlled trials (RCTs) have long been considered the gold standard of research methodology (Borenstein et al., 2009; Littell, Corcoran, & Pillai, 2008; West, 2009). Although RCTs are not as popular in the organizational sciences, they have been used extensively for the evaluation of

therapeutic interventions in the biological, medical, and health sciences (Peto et al., 1976; Yusuf, Collins, & Peto, 1984). Because of the methodological rigor associated with RCTs, an important question that has been raised is what to conclude when there is a discrepancy between results of a meta-analysis and a large sample RCT. Researchers have expressed this concern because meta-analyses and large-sample RCTs have yielded divergent results, which leads researchers to question the validity of meta-analysis (Ioannidis, Cappelleri, & Lau, 1998). Moreover, these discrepancies have resulted in strong criticisms against “the arm-chair research done by meta-analyzers” (Kelly, 1997, p. 1182). Thus, the MUL is that when a discrepancy occurs between meta-analytic and RCT results, the attribution is that there must be a problem with the meta-analysis.

Consider the following examples to illustrate this MUL. Flather, Farkouh, Pogue, and Yusuf (1997) reviewed studies reporting meta-analyses and RCTs of the effects of magnesium and thrombolytic therapy as two types of therapeutic treatment for acute myocardial infarction. After finding discrepancies in results obtained using meta-analysis versus RCTs, Flather et al. (1997) concluded that “the main reasons for this disagreement are publication bias and the small size of the meta-analysis” (p. 576). Similarly, after describing discrepancies between meta-analytic and RCT results, Cappelleri et al. (1996) concluded that “clinicians and other decision makers should realize that publication bias, study protocol, and variability in the control rate of events in different trials may underlie discordant results” (p. 1337). As a third example, in a *New England Journal of Medicine* article, LeLorier, Gregoire, Benhaddad, Lapierre, & Derderian (1997) compared the results of 12 RCTs that included at least 1,000 patients and were published in 4 major medical journals with results of 19 published meta-analyses on the same topics. After reporting that meta-analytic results could not predict RCT results accurately 35% of the time, the authors concluded that the likely reasons were the following problems with the meta-analyses: (a) publication bias (i.e., “a meta-analysis that excluded unpublished studies or did not locate and include them would thus be more likely to have a false positive result,” p. 540) and (b) heterogeneity of studies included in the meta-analyses (i.e., “the heterogeneity of the trials included in the meta-analysis may partially account for divergence of this type,” p. 540). LeLorier et al. (1997) provided a clear answer to the question of which results one should trust in the presence of discrepancies, “How should clinicians use meta-analyses, given that systematic comparison with randomized clinical trials shows that they have poor predictive ability? Most will agree that if a large, well-done randomized trial has been conducted, practice guidelines should be strongly influenced by its results” (p. 541).

With respect to this MUL, there are several valid reasons why meta-analytic results may differ from RCT results, and these reasons are not necessarily related to a deficiency, inaccuracy, or errors in the meta-analysis (Borenstein et al., 2009). Suppose, for example, that a meta-analyst synthesizes results from several RCTs to examine the overall effect of an independent variable (e.g., type of training) on a dependent variable (e.g., trainees’ reactions to the type of training received). The meta-analysis can produce a mean effect-size estimate that summarizes the overall magnitude of the relationship between the independent and dependent variable across the RCTs (Petticrew & Roberts, 2006). It is quite possible that this meta-analytically derived mean effect-size estimate does not match the value of any of the RCT-level (i.e., primary study-level) effect-size estimates. Is this discrepancy between results of a meta-analysis and RCTs problematic? Does it suggest that meta-analytic techniques are defective? As we explain next, the answer to each of these questions is no.

Considering that the study-level effect-size estimates can vary across RCTs, it would not be alarming if the meta-analytically derived mean effect-size estimate did not match the value of any of the RCT-level effect-size estimates. Hence, the concern should not be whether there is a discrepancy between effect-size estimates produced from a meta-analysis versus RCTs, but rather whether and to what extent there is variability in study-level effect-size estimates across the RCTs (Borenstein et al., 2009). Variability in effect-size estimates across RCTs may be due to moderator variables or, alternatively, to artifactual sources such as sampling error (Le, Schmidt, & Putka,

2009). An important goal of a meta-analysis that synthesizes results from RCTs should thus be to explain the variability in RCT-level effect-size estimates.

Although some researchers question the validity of meta-analytic results when they differ from RCT results, it is important to take a balanced perspective on this discrepancy issue. RCTs have long been considered the gold standard of research methodology, but they too have limitations. For example, a randomized experiment can turn into a “broken” design because of attrition of study participants or their noncompliance with an assigned treatment condition (West, 2009). Thus, both meta-analysis and RCTs each have methodological strengths and weaknesses. As noted by Yusuf (1997), “meta-analyses are not replacements for large trials nor are large trials replacements for meta-analyses” (p. 600).

In sum, the MUL is that a discrepancy between results of a meta-analysis and RCTs suggests that the meta-analysis is defective. The kernel of truth value is that, under some conditions, it may be that differences are due to problems in the design and/or execution of a meta-analysis such as the inclusion of poor-quality studies (see MUL #2). However, the misunderstanding is that, in most cases, differences between results of a meta-analysis and RCTs are to be expected because the meta-analytic estimate is an average of the primary-level effect sizes. By definition, the mean across-study effect size is the best representative of the set of primary-level effect sizes and, at the same time, will differ from some of them (unless there is no across-study variance in primary-level effect sizes). In terms of meta-analytic practice, the recommendation is not to focus on discrepancies between results of meta-analyses and RCTs. Instead, the focus should be on using meta-analysis to explain across-study variability in RCTs’ effect-size estimates (e.g., due to moderator variables or, alternatively, to artifactual sources of variance such as sampling error).

MUL #7: Meta-Analytic Technical Refinements Lead to Important Scientific and Practical Advancements

As is the case with most methodological approaches, meta-analysis is subject to ongoing refinements and improvements. Moreover, there is a belief that technical refinements produce important theory advancements, lead to important changes in substantive conclusions, and also lead to substantively revised and improved recommendations for practice (Aguinis, 2001; Aguinis & Whitehead, 1997; Hunter, Schmidt, & Le, 2006; Le, Oh, Shaffer, & Schmidt, 2007; Oh, Schmidt, Shaffer, & Le, 2008; Schmidt, Oh, & Le, 2006). For example, consider refinements addressing range restriction, which occurs when a sample is subject to selection (e.g., only the subset of highest performing firms from the population choose to return a survey, or a validity coefficient is computed including students whose college admissions scores were above a certain cutoff). Aguinis and Whitehead (1997) concluded that “in the presence of indirect range restriction, variability across study-level r s can be underestimated by as much as 8.50%” (p. 537). Based on this result, Aguinis and Whitehead (1997) recommended that future meta-analyses correct for the effects of indirect range restriction. More recent research regarding adjustments due to newly developed indirect range restriction corrections has led to the conclusion that “Meta-analysis results produced when one corrects for indirect range restriction produce larger mean values in comparison with those produced by application of the inappropriate correction for direct range restriction” (Hunter et al., 2006, p. 608). In spite of the conclusion that refinements lead to “larger mean values,” an improvement in a meta-analytically derived mean correlation, for example, from .30 to .32 may seem impressive if this is described as a “7% improvement” in the relationship between a predictor and an outcome of interest (Oh et al., 2008). However, as noted by Aguinis, Dalton, et al. (in press), “this type of ‘improvement’ means that we now explain 10% of variance in an outcome rather than 9%. Although such a small increase may be practically significant in a handful of contexts (i.e., use of the Graduate Management Admission Test [GMAT] with thousands of graduate student applicants worldwide), it is

hardly large enough to conclude that we now explain the outcome so much better that the practical usefulness of a theory is substantially greater.” In spite of the conclusion of Aguinis, Dalton, et al. that refinements in meta-analysis methods generally may not result in radical revisions of theories, there may be situations in which changes brought about by such refinements can potentially be significant, particularly in areas where small changes in an effect size can have important practical implications (Aguinis, Werner, et al., 2010).

Aguinis, Dalton, et al. (in press) examined the extent to which 21 methodological choices and judgment calls have an important effect on substantive conclusions in published meta-analyses. The choices and judgment calls examined ranged from the stage of primary-level study retrieval (e.g., Were any studies eliminated and why?) to the data-analysis stage (i.e., Was a fixed-effect or a random-effect model used?) and to the results-reporting stage (i.e., Was a file drawer analysis conducted?). Many of these choices and judgment calls are directly related to meta-analytic technical refinements. The review by Aguinis, Dalton, et al. involved a content analysis of 196 meta-analyses including 5,581 effect-size estimates published in *AMJ*, *JAP*, *JOM*, *PPsych*, and *SMJ* from January 1982 through August 2009. Specifically related to our earlier discussion on range restriction, Aguinis, Dalton, et al. (in press; Table 1) calculated that authors acknowledged that about 12% of meta-analytically derived effect sizes were affected by range restriction on the independent variable and approximately 10% of analytically derived effect sizes were affected by range restriction on the dependent variable. Overall, results suggest that the various methodological choices and judgment calls involved in the conduct of a meta-analysis have little impact on the obtained effect sizes. Accordingly, these results based on actual published meta-analyses and not on simulated data casts doubt on previous severe warnings, primarily based on selective case studies, that judgment calls have an important impact on substantive conclusions. As noted by Murphy (2003) regarding validity generalization (VG; i.e., psychometric meta-analysis; Hunter & Schmidt, 2004), in particular, but equally applicable to any type of meta-analytic procedure, “a central weakness of most tests of the accuracy of VG estimates is the gap between the assumptions needed to develop and test these models and the actual process by which validity results are produced and generated . . . assumptions of normality, random sampling, independence, and so forth are routinely violated in most studies in the behavioral and social sciences” (p. 17).

At the risk of some level of meta-analytic Luddism, immediate improvement, at margin, of meta-analytic applications is more likely to be realized by increased attention to other threats to its sound execution. Bobko and Roth (2008) provided a compelling perspective on this point. They noted that researchers have appropriately accepted meta-analytic applications but may have focused too much energy on its refinements. They suggest, no matter how good such refinements may be, and they concede that many are inspired, “meta-analysis is not necessarily much better than the primary data, thinking, and theory that go into it at the beginning” (p. 115).

In sum, the MUL is that meta-analytic technical refinements have important consequences for theory and application. The kernel of truth value is that such refinements lead to improved accuracy and estimation that may be meaningful in certain contexts only. For example, an improvement in a validity coefficient from .15 to .19 may have important practical implications in the context of hiring thousands of job applicants. In such a context including so many job applicants, using a more valid test, albeit just slightly more valid, may lead to practical improvements in terms of false positive and false negative selection decisions (Aguinis & Smith, 2007). However, in most organizational science research, such improvements do not have a meaningful and substantive impact on theory or practice. In terms of recommendations for meta-analytic practice, researchers should always use the most accurate estimation procedures available. To be clear, we are not advocating that new refinements should be ignored. In fact, “a hallmark of a science is continuous improvement in accuracy of measurements and estimations of theoretically important values” (Aguinis, 2001, p. 587). Moreover, we agree with Schmidt and Hunter (2003) who noted that “even if estimates are quite accurate,

it is always desirable to make them more accurate if possible” (p. 41). However, the belief that such refinements lead to, in most cases, important and substantive theoretical advancements and improved application is typically not justified.

Discussion

Consider the legend of the philosopher’s stone. This stone, the enabler of the alchemist’s obsession, could transform base metals into gold, serve as the elixir of life, and bestow immortality upon those who possessed it. Meta-analysis, as enabling as it is, does not have this character. It has been astutely suggested that “the goal of any empirical science is to pursue the construction of a cumulative base of knowledge upon which the future of the science may be built” (Curran, 2009, p. 77). For us, meta-analysis is an extraordinary enablement to facilitate that journey. Moreover, we are unabashed advocates of meta-analysis. Having said that, while meta-analysis is “extraordinary,” it is far from perfect. As noted by Cooper and Hedges (2009), perfect primary studies do not exist; perfect syntheses do not exist. Years ago, T. S. Eliot in the opening stanza from “Choruses from the Rock” (1934) asked the question, “Where is the knowledge we have lost in the information?” Meta-analysis has the promise to address this question, to retrieve knowledge from many single sources of information. For now, meta-analysis is the definitive means of summarizing a body of empirical research. Alas, as noted, it will not do so perfectly, but it provides “a model of the kind of paradigm shift that is possible” (Shrout, 2009, p. 180). In addition to its ability to synthesize an existing literature, meta-analysis provides an additional paradigm shift that is largely without comparison and underscores the necessity of full disclosure in its applications.

Disclosure and Replicability

Any research should be described in sufficient detail so that it can be replicated. Heraclitus (535-475 BCE), presumably in a different context, provided the counterperspective that “[n]o man ever steps in the same river twice, for it’s not the same river and he’s not the same man.” With meta-analysis, however, it is possible to put one’s toe back in the metaphorical water. In that spirit, we suggest that meta-analysis is actually the near-perfect vehicle for disclosure and replicability (Dalton & Dalton, 2005; Eden, 2002). In fact, all meta-analyses should be subject to strict replication.

For all meta-analyses, the actual studies from which the data were derived should be listed within the article. Accordingly, any potential meta-analyst can verify and replicate any of the input data including the effect sizes, number of studies, and samples from which those data were derived. The meta-analyst should also be able to review information on reliability estimates, range restriction, data inclusion/exclusion criteria, and other assumptions on which the meta-analysis relied. Also available should be the exact ranges on which confidence and credibility intervals were based. Such information should be available for direct effects as well as moderator analyses. Given this information, it would be possible to reanalyze meta-analysis data with alternative assumptions that were not relied on for the initial synthesis. Beyond that, the relevant data can be reanalyzed to include studies that were not identified, or not included, or those that did not exist at the time of the initial meta-analysis. This potential for replication is unprecedented. Rothstein, McDaniel, and Borenstein (2002, p. 538) commented that “[e]xplicitness and transparency of procedures are hallmarks of a properly conducted meta-analysis.” The potential subjection of meta-analysis to this extraordinary level of scrutiny reinforces its credibility and, thus, its application to policy and practice (Lipsey & Wilson, 2001). However, for meta-analysis to realize this potential, reporting practices should change so that sufficient information is provided in all published meta-analyses to allow for such

replication (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008).

Closing Comments

Despite the many advantages of meta-analysis, we echo sentiment by Bobko and Stone-Romero (1998) that meta-analysis is not a panacea and is not—and has never been—a substitute for primary-level research (e.g., Cooper & Hedges, 2009). Meta-analytic and primary-level research are, however, complementary. Many years ago, Feldman (1971, p. 100) provided a prescient observation, “A good integration, at the same time that it shows how much is known in an area, also shows how little is known. It sums up, but does not end. In this sense, it is only a beginning.” Very well said then, as now.

Authors' Note

A previous version of this manuscript was presented at the meetings of the Society for Industrial and Organizational Psychology, Atlanta, Georgia, April 2010.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

- Aguinis, H. (2001). Estimation of sampling variance of correlations in meta-analysis. *Personnel Psychology, 54*, 569-590.
- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: Guilford Press.
- Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90*, 94-107.
- Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology, 95*, 648-680.
- Aguinis, H., Dalton, D. A., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (in press). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*.
- Aguinis, H., Gottfredson, R. K., & Wright, T. A. (in press). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior*.
- Aguinis, H., & Harden, E. E. (2009). Cautionary note on conveniently dismissing χ^2 goodness-of-fit test results: Implications for strategic management research. In D. D. Bergh & D. J. Ketchen (Eds.), *Research methodology in strategy and management* (Vol. 5, pp. 111-120). Howard House, UK: Emerald Group Publishing.
- Aguinis, H., Mazurkiewicz, M. D., & Heggstad, E. D. (2009). Using Web-based frame-of-reference training to decrease biases in personality-based job analysis: An experimental field study. *Personnel Psychology, 62*, 405-438.
- Aguinis, H., & Pierce, C. A. (1998). Testing moderator variable hypotheses meta-analytically. *Journal of Management, 24*, 577-592.
- 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.

- Aguinis, H., Pierce, C. A., & Culpepper, S. A. (2009). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods, 12*, 623-652.
- Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology, 60*, 165-199.
- Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82*, 192-206.
- Aguinis, H., Sturman, M. C., & Pierce, C. A. (2008). Comparison of three-meta-analytic procedures for estimating moderating effects of categorical variables. *Organizational Research Methods, 11*, 9-34.
- Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods, 13*, 515-539.
- Aguinis, H., & Whitehead, R. (1997). Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. *Journal of Applied Psychology, 82*, 528-538.
- Almekhlafi, M. A., Wilton, S. B., Rabi, D. M., Ghali, W. A., Lorenzetti, D. L., & Hill, M. D. (2009). Recurrent cerebral ischemia in medically treated patent foramen ovale: A meta-analysis. *Neurology, 73*, 89-97.
- APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63*, 839-851.
- Becker, B. J. (2005). 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. 111-125). West Sussex, England: John Wiley.
- Bedeian, A. G., Van Fleet, D. D., & Hyman, H. H. (2009a). Scientific achievement and editorial board membership. *Organizational Research Methods, 12*, 211-238.
- Bedeian, A. G., Van Fleet, D. D., & Hyman, H. H. (2009b). "Circle the wagons and defend the faith:" Slicing and dicing the data. *Organizational Research Methods, 12*, 276-295.
- Begg, C. B. (1994). Publication bias. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-409). New York, NY: Russell Sage.
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088-1101.
- Bergh, D. D., Aguinis, H., Hanke, R., & Perry, J. (2010, August). *Resource allocation strategies, shareholder wealth, and firm financial performance*. Paper presented at the meetings of the Academy of Management, Montréal, Canada.
- Bobko, P., & Roth, P. L. (2003). Meta-analysis and validity generalization as research tools: Issues of sample bias and degrees of mis-specification. In K. Murphy (Ed.), *Validity generalization: A critical review* (pp. 67-90). Mahwah, NJ: Lawrence Erlbaum.
- Bobko, P., & Roth, P. L. (2008). Psychometric accuracy and (the continuing need for) quality thinking in meta-analysis. *Organizational Research Methods, 11*, 114-126.
- Bobko, P., & Stone-Romero, E. F. (1998). Meta-analysis may be another useful research tool, but it is not a panacea. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 16, pp. 359-397). Greenwich, CT: JAI Press.
- Bonett, D. G. (2008). Meta-analytic interval estimation for bivariate correlations. *Psychological Methods, 13*, 173-189.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, England: John Wiley.
- Brind, J., Chinchilli, V. M., Severs, W. B., & Summy-Long, J. (1996). Induced abortion as an independent risk factor for breast cancer: Comprehensive review and meta-analysis. *Journal of Epidemiology and Community Health, 50*, 481-496.

- Briner, R. B., Denyer, D., & Rousseau, D. M. (2009). Evidence-based management: Concept cleanup time? *Academy of Management Perspectives*, 23, 19-32.
- Cappelleri, J. C., Ioannidis, J. P. A., Schmid, C. H., de Ferranti, D. D., Aubert, M., Chalmers, T. C., & Lau, J. (1996). Large trials vs meta-analyses of small trials: How do their results compare? *Journal of the American Medical Association*, 276, 1332-1338.
- Carlson, K. D., & Ji, F. X. (2009, April). *Citing, reporting, interpreting, and building on meta-analytic findings*. Paper presented at the meetings of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, 93, 1062-1081.
- Chan, A., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *Journal of the American Medical Association*, 291, 2457-2465.
- Cheung, M. W. L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods*, 13, 182-202.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, 8, 243-253.
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678-707.
- Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Los Angeles, CA: Sage.
- Cooper, H., & Hedges, L. V. (2009). Potentials and limitations. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 561-572). New York, NY: Russell Sage.
- Cooper, H., Hedges, L. V., & Valentine, J.C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. (2nd ed.). New York, NY: Russell Sage.
- Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods*, 6, 415-439.
- Curran, P. J. (2009). The seemingly quixotic pursuit of a cumulative psychological science. *Psychological Methods*, 14, 77-80.
- Dalton, D. R., & Dalton, C. M. (2005). Strategic management as a special case for meta-analysis. In D. J. Ketchen & D. D. Bergh (Eds.), *Research methodology in strategy and management* (Vol. 3, pp. 31-63). New York, NY: Elsevier.
- Dalton, D. R., & Dalton, C. M. (2008). Meta-analyses: Some very good steps toward a bit longer journey. *Organizational Research Methods*, 11, 127-147.
- Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 11-33). West Sussex, England: John Wiley.
- Dieckmann, N. F., Malle, B. F., & Bodner, T. E. (2009). An empirical assessment of meta-analytic practice. *Review of General Psychology*, 13, 101-115.
- Duval, S. (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. 127-144). West Sussex, England: John Wiley.
- Duval, S. J., & Tweedie, R. L. (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. 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, 455-463.

- Eden, D. (2002). Replication, meta-analysis, scientific progress, and AMJ's publication policy. *Academy of Management Journal*, 45, 841-846.
- 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.
- Egger, M., Smith, G. D., & Altman, D. G. (2001). *Systematic reviews in health care: Meta-analysis in context* (2nd ed.). London, UK: BMJ.
- Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, 33, 517.
- Eysenck, H. J. (1984). Meta-analysis: An abuse of research integration. *Journal of Special Education*, 18, 41-59.
- Feldman, K. A. (1971). Using the work of others: Some observations on reviewing and integrating. *Sociology of Education*, 44, 86-102.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161-180.
- Flather, M. D., Farkouh, M. E., Pogue, J. M., & Yusuf, S. (1997). Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Controlled Clinical Trials*, 18, 568-579.
- Furlow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods*, 10, 227-254.
- Geyskens, I., Krishnan, R., Steenkamp, J.-B. E. M., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management*, 35, 393-419.
- Gilliland, S. W., & Cortina, J. M. (1997). Reviewer and editor decision making in the journal review process. *Personnel Psychology*, 50, 427-452.
- Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of research in education*, 5, 351-379.
- Glass, G. V. (1978). In defense of generalization. *Behavioral and Brain Sciences*, 1, 394-395.
- González-Benito, J., Aguinis, H., Boyd, B. K., & Suárez-González, I. (2010). *Coming to consensus on strategic consensus: A mediated moderation model of consensus and performance*. Manuscript submitted for publication.
- Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33, 920-934.
- Hall, J. A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs*, 58, 437-448.
- Halpern, S. D., & Berlin, J. A. (2005). Beyond conventional publication bias: Other determinants of data suppression. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 303-317). West Sussex, England: John Wiley.
- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, 36, 865-891.
- Härlein, J., Dassen, T., Halfens, R. J. G., & Heinze, C. (2009). Fall risk factors in older people with dementia or cognitive impairment: A systematic review. *Journal of Advanced Nursing*, 65, 922-933.
- Hartung, J., Knapp, G., & Sinha, B. K. (2008). *Statistical meta-analysis with applications*. Hoboken, NJ: John Wiley.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426-445.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed and random effects models in meta-analysis. *Psychological Methods*, 3, 486-504.
- Higgins, J. P. T., & Green S. (Eds.). (2009). *Cochrane handbook for systematic reviews of interventions version 5.0.2* [updated September 2009]. The Cochrane Collaboration. Retrieved from www.cochrane-handbook.org
- Hitt, M. A. (2009). Editorial judgments, quality scholarship, and the Academy of Management's journals. *Organizational Research Methods*, 12, 253-258.

- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). New York, NY: Academic Press.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*, 594-612.
- Ioannidis, J. P., Cappelleri, J. C., & Lau, J. (1998). Issues in comparisons between meta-analyses and large trials. *Journal of the American Medical Association, 279*, 1089-1093.
- Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282*, 1054-1060.
- Kastner, M., & Straus, S. E. (2008). Clinical decision support tools for osteoporosis disease management: A systematic review of randomized controlled trials. *Journal of General Internal Medicine, 23*, 2095-2105.
- Kelly, C. P. (1997). Meta-analysis. *Lancet, 350*, 1182.
- Klimoski, R. (2009). Begin the journey with the end in mind. *Organizational Research Methods, 12*, 239-252.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*, 202-220.
- Lance, C. E., & Vandenberg, R. J. (2009). *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences*. New York, NY: Routledge.
- Le, H., Oh, I., Shaffer, J. A., & Schmidt, F. L. (2007). Implications of methodological advances for the practice of personnel selection: How practitioners benefit from meta-analysis. *Academy of Management Perspectives, 21*, 6-15.
- Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organizational Research Methods, 12*, 165-200.
- LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre, J., & Derderian, F. (1997). Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine, 337*, 536-542.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford, England: Oxford University Press.
- McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology, 59*, 927-953.
- Muncer, S., Taylor, S., & Craigie, M. (2002). Power dressing and meta-analysis: Incorporating power analysis into meta-analysis. *Journal of Advanced Nursing, 38*, 274-280.
- Murphy, K. R. (2003). The logic of validity generalization. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 1-29). Mahwah, NJ: Lawrence Erlbaum.
- Oh, I., Schmidt, F. L., Shaffer, J. A., & Le, H. (2008). The Graduate Management Admission Test (GMAT) is even more valid than we thought: A new development in meta-analysis and its implications for the validity of the GMAT. *Academy of Management Learning and Education, 7*, 563-570.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*, 4544-4562.
- Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., . . . Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer, 34*, 585-612.
- Petticrew, M. (2001). Systematic reviews from astronomy to zoology: Myths and misconceptions. *British Medical Journal, 322*, 98-101.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences*. Malden, MA: Blackwell.
- Pfeffer, J., & Sutton, R. I. (2006). Evidence based management. *Harvard Business Review, 84*, 62-74.
- Pigott, T. D. (2009). Handling missing data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 399-416). New York, NY: Russell Sage.

- Podsakoff, P. M., & Dalton, D. R. (1987). Research methodology in organizational studies. *Journal of Management*, *13*, 419-441.
- Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*, 1-12.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Some recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59-82.
- Rothstein, H. R., McDaniel, M. A., & Borenstein, M. (2002). Meta-analysis: A review of quantitative cumulation methods. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 534-570). San Francisco, CA: Jossey-Bass.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). 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). West Sussex, England: John Wiley.
- Rousseau, D. M., 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.
- Russell, C. J., Settoon, R. P., McGrath, R. N., Blanton, A. E., Kidwell, R. E., Lohrke, F. T., . . . Danforth, G. W. (1994). Investigator characteristics as moderators of personnel selection research: A meta-analysis. *Journal of Applied Psychology*, *79*, 163-170.
- Rynes, S. L., Bartunek, J. M., & Daft, R. L. (2001). Across the great divide: Knowledge creation and transfer between practitioners and academics. *Academy of Management Journal*, *44*, 340-355.
- Sackett, P. R., Harris, M. M., & Orr, J. M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to Type I error. *Journal of Applied Psychology*, *71*, 302-310.
- Sagie, A., & Koslowsky, M. (1993). Detecting moderators with meta-analyses: An evaluation and comparison of techniques. *Personnel Psychology*, *46*, 629-640.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173-1181.
- Schmidt, F. L. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods*, *11*, 96-113.
- Schmidt, F., & Hunter, J. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975-2001. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 31-65). Mahwah, NJ: Lawrence Erlbaum.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsh, H. R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, *38*, 697-798.
- Schmidt, F. L., Oh, I.-S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology*, *59*, 181-305.
- Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hogrefe & Huber Publishing.
- Shadish, W. R., Cook, D. T., & Campbell, T. D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, *65*, 98-109.
- Shieh, G. (2009). Detecting interaction effects in moderated multiple regression with continuous variables: Power and sample size considerations. *Organizational Research Methods*, *12*, 510-528.
- Shrout, P. E. (2009). Short and long views of integrative data analysis: Comments on the contributions to the special issue. *Psychological Methods*, *14*, 177-181.

- Sood, A. K. (1991). Cigarette smoking and cervical cancer: Meta-analysis and critical review of recent studies. *American Journal of Preventive Medicine, 7*, 208-213.
- Stangl, D. K., & Berry, D. A. (2000). *Meta-analysis in medicine and health policy*. New York, NY: Marcel Dekker.
- Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic publications. *Organization Science, 16*, 180-200.
- Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 75-98). West Sussex, England: John 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). West Sussex, England: John Wiley.
- Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435-452). New York, NY: Russell Sage.
- Sutton, A. J., Abrams, K. R., Ades, A. S., Copper, N. J., & Welton, N. J. (2009). *Evidence synthesis for decision making in healthcare*. Chichester, UK: Wiley.
- 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.
- Tsui, A. S., & Hollenbeck, J. R. (2009). Successful authors and effective reviewers: Balancing supply and demand in the organizational sciences. *Organizational Research Methods, 12*, 259-275.
- Valentine, J. C. (2009). Judging the quality of primary research. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 129-146). New York, NY: Russell Sage.
- Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods, 13*, 130-149.
- Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods, 9*, 194-201.
- Vandenberg, R. J., & Grelle, D. M. (2009). Alternative model specifications in structural equation modeling: Facts, fictions, and truth. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences* (pp. 165-191). New York, NY: Routledge.
- Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*, 865-885.
- Wanous, J. P., Sullivan, S. E., & Malinak, J. (1989). The role of judgment calls in meta-analysis. *Journal of Applied Psychology, 74*, 259-264.
- Weed, D. L. (2000). Interpreting epidemiological evidence: How meta-analysis and causal inference methods are related. *International Journal of Epidemiology, 29*, 387-390.
- West, S. G. (2009). Alternatives to randomized experiments. *Current Directions in Psychological Science, 18*, 299-304.
- Yusuf, C. (1997). Meta-analysis of randomized trials: Looking back and looking ahead. *Controlled Clinical Trials, 18*, 594-601.
- Yusuf, S., Collins, R., & Peto, R. (1984). Why do we need some large, simple randomized trials? *Statistics in Medicine, 3*, 409-422.

Bios

Herman Aguinis is the dean's research professor, a professor of organizational behavior and human resources, and the director of the Institute for Global Organizational Effectiveness in the Kelley School of Business,

Indiana University. His research interests span several human resource management, organizational behavior, and research methods and analysis topics.

Charles A. Pierce is a professor of management in the Department of Management, Fogelman College of Business and Economics, University of Memphis. His research interests include workplace romance, sexual harassment, and organizational research methods.

Frank A. Bosco is a PhD candidate in the Department of Management, Fogelman College of Business and Economics, University of Memphis. His research interests include ethical decision making, staffing, and organizational research methods.

Dan R. Dalton is the founding director of the Institute for Corporate Governance, Dean Emeritus, and the Harold A. Poling Chair of Strategic Management in the Kelley School of Business, Indiana University. A Fellow of the Academy of Management, his research focuses on corporate governance and research methods and analysis.

Catherine M. Dalton holds the David H. Jacobs Chair of Strategic Management in the Kelley School of Business, Indiana University. She has published in the areas of corporate governance and strategic leadership, business ethics, and research methods.