Journal of Leadership & Organizational Studies
2016, Vol. 23(1) 66–81

DOI: 10.1177/1548051815614321
jlo.sagepub.com

An Assessment of the Magnitude of Effect Sizes: Evidence From 30 Years of Meta-Analysis in Management

Ted A. Paterson[1], P.D. Harms[2], Piers Steel[3], and Marcus Credé[4]

Abstract

This study compiles information from more than 250 meta-analyses conducted over the past 30 years to assess the magnitude of reported effect sizes in the organizational behavior (OB)/human resources (HR) literatures. Our analysis revealed an average uncorrected effect of $r = .227$ and an average corrected effect of $\rho = .278$ ($SD\rho = .140$). Based on the distribution of effect sizes we report, Cohen's effect size benchmarks are not appropriate for use in OB/HR research as they overestimate the actual breakpoints between small, medium, and large effects. We also assessed the average statistical power reported in meta-analytic conclusions and found substantial evidence that the majority of primary studies in the management literature are statistically underpowered. Finally, we investigated the impact of the file drawer problem in meta-analyses and our findings indicate that the file drawer problem is not a significant concern for meta-analysts. We conclude by discussing various implications of this study for OB/HR researchers.

Keywords

effect size, meta-analysis, statistical power, file drawer problem

> Integration is one of the primary objectives of any scientist. It is his dream to see his field someday begin to have order. He gains as much satisfaction from seeing that someone has integrated hitherto seemingly discrete information as the reader might receive from reading a novel, seeing an exciting movie, or going to the World Series.
>
> —Argyris (1957, p. 3)

Despite a long history of relying almost exclusively on significance tests in organizational research, there is a renewed emphasis on effect size reporting in the organizational behavior (OB) and human resources (HR) literatures (American Psychological Association, 2001; Campion, 1993; Zedeck, 2003). This development is a positive one because routine reporting of effect sizes has the potential to improve the assessment of "practical significance" (Kirk, 1996), statistical power analysis (Cohen, 1988), meta-analysis, comparison of effects across studies, among other key aspects of micro-oriented management research. However, given the sheer size of the OB and HR literatures, identification and evaluation of all relevant prior studies in order to aid with the interpretation of effect sizes in current studies would be nearly impossible. Thus, a high level review or aggregation is needed that summarizes the magnitude and distribution of effect sizes in OB and HR fields generally and with respect to the various sub fields within these major areas. The purpose of this article is to provide such a review.

The OB/HR literatures have long been characterized by attempts to synthesize different research findings in a particular domain of inquiry. Early attempts at such synthesis, typically narrative in nature (e.g., leadership literature, Bass, 1954; Jenkins, 1947; Smith & Krueger, 1933; Stogdill, 1948), helped researchers to take stock of progress in a domain, compare results across studies, and recognize understudied areas. More recently, narrative reviews have been supplemented by quantitative summaries of research—typically taking the form of meta-analyses of effect sizes—that synthesize the findings on a particular relationship found across multiple studies to arrive at three important estimates: (1) an estimate of the average population effect size, (2) an estimate of how much of the observed variability in effect sizes across studies could be attributed to sampling error and other study artifacts, and hence (3) an estimate of the degree to which the strength of the relationship may be moderated by unexamined moderators. As

[1]Oregon State University, Corvallis, OR, USA
[2]University of Alabama, Tuscaloosa, AL, USA
[3]University of Calgary, Calgary, Alberta, Canada
[4]Iowa State University, Ames, IA, USA

Corresponding Author:
Ted A. Paterson, College of Business, Oregon State University, 352 Austin Hall, Corvallis, OR 97331, USA.
Email: ted.paterson@oregonstate.edu

such, meta-analyses have been particularly valuable in areas of research characterized by seemingly disparate research findings. That is, meta-analyses have allowed researchers to either conclude that variations in observed correlations are simply due to study artifacts or that such variations are likely the result of moderators (known or unknown). For example, the influential review of the job satisfaction—job performance literature by Judge, Thoresen, Bono, and Patton (2001) suggested that the relationship (ρ = .30) is substantially moderated by unknown variables. Similarly, Kluger and DeNisi's (1996) review of the impact of feedback interventions on performance was able to identify a number of feedback characteristics that either augmented or attenuated the effect of feedback interventions on performance but also concluded that task characteristics were likely moderators of the effectiveness of feedback interventions but that the precise nature of these task characteristics were not well understood. Therefore meta-analyses serve not only as influential summaries of a field but can also serve to suggest new areas of inquiry within a particular research domain or highlight how specific moderators can explain apparently disparate research findings.

In the same manner in which meta-analytic reviews of specific empirical literatures provide useful summaries of past work in an area and suggest potentially fruitful areas of future inquiry, so might an examination of the meta-analyses of the broader OB/HR literatures provide insights into past research (and past research practices), and suggest where micro-oriented management researchers might best focus their attention and resources when deciding on future areas of inquiry. Such examinations have previously been undertaken by other researchers. For example, Richard, Bond, and Stokes-Zoota (2003) conducted a review of social psychological research, although their review specifically excluded studies that fell within the industrial/organizational psychology realm. Other reviews of meta-analyses include Hunter and Hirsh's (1987) narrative review of meta-analyses from OB research; Steiner, Lane, Dobbins, Schnur, and McConnell's (1991) empirical assessment of variability in meta-analysis techniques; and Tett, Meyer, and Roese's (1994) follow-up of Hunter and Hirsh's (1987) work. More recently, Aguinis, Dalton, Bosco, Pierce, and Dalton (2011) conducted a quantitative review of how decisions in the meta-analytic process affected reported results. Our article aims to supplement these reviews by conducting a quantitative review of meta-analytic publications of OB/HR research with a specific focus on the magnitude and variability of the effect sizes reported.

In the sections that follow, we describe the importance of effect size magnitude, the difficulties associated with comparing effect sizes, and issues with statistical power in micro-oriented management research. Next, we explore evidence regarding the degree to which OB/HR researchers should be concerned with the file drawer problem. We then describe our quantitative review of 30 years of meta-analysis in the organizational sciences and how it contributes to various ongoing debates regarding the magnitude and variability of effect sizes and factors which affect each. Finally, we describe our findings and implications for researchers.

Magnitude of Effect Sizes

For nearly a century, null hypothesis significance testing (NHST) has been a central part of the social sciences. However, NHST has been the center of controversies and criticisms for equally as long (Kirk, 1996). Chief among these criticisms are the claims that NHST (1) does not tell researchers what they want to know, (2) is a trivial exercise and (3) it turns a continuum of uncertainty into a dichotomous decision (Aguinis et al., 2010; Cohen, 1994; Kirk, 1996). Even with these and other criticisms, NHST has been ritualized (Cohen, 1994) into modern scientific practices despite being described as the "statistical folkways of a more primitive past" (Rozeboom, 1960, p. 417). Given the fact that NHST has been shrouded in controversy since its inception, it is not surprising to learn that calls to supplement or replace the NHST have been long-standing. Fisher (1925) was perhaps the first to recommend that researchers supplement the significance test by reporting the correlation ratio (Kirk, 1996). Yates (1951) echoed this sentiment when he stated that "scientific research workers pay undue attention to the results of the tests of significance they perform on their data . . . and too little to the estimates of the magnitude of the effects they are estimating" (p. 32). Because NHST combines the effect size and the sample size, the reporting of this test alone does not give the reader an indication of the magnitude of the effect. The effect size provides consumers of research with an estimate of the strength of the relationship between the independent and dependent variables and is not conflated with the sample size, as is the NHST. Notwithstanding the clear advantages of reporting the effect size in conjunction with the NHST, this practice has not found widespread use despite the calls from numerous scholars over multiple decades (e.g., Coe, 2002; Cohen, 1994; Glass & Hakstian, 1969; Peters & Van Voorhis, 1940).

Of all of these voices calling attention to the shortcomings of NHST and the need for effect size reporting, Cohen's (1990, 1994) has been the most influential. As Kirk (1996) explains, Cohen's work was especially impactful because his effect size (*d*) came with guidelines for interpreting the magnitude of the effect. Why is this of concern to OB/HR researchers? The comparative size of the effects in a field or area of inquiry has bearing on the scientific status of the field (Rosenthal, 1994) and the predictive ability of its theories (Bacharach, 1989). However, it is also important to point out that a small effect size is not necessarily indicative of a meaningless finding (Prentice & Miller, 1992). As

Rosenthal (1994) stated, seemingly small effects can have significant real-world consequences. However, all else being equal, larger effects tend to have more important consequences for both researchers and practitioners.

According to Cohen, "many" effects in the social sciences are small ($r = .10$) in size. Others have reported varying estimates of the average effect size in management research. Mazen, Graf, Kellogg, and Hemmasi (1987) concluded that the average effect size in their sample of 12 meta-analyses was $d = .39$ ($|r| = .19$). Aguinis et al. (2011) reported an average corrected effect size of $|r| = .261$ ($d = .54$) from their sample of 196 meta-analyses, an effect that is substantially larger than that reported by Mazen et al. (1987). More recently, a compilation of effect sizes from two premier applied psychology journals reported that the average effect size was $|r| = .16$ (Bosco, Aguinis, Singh, Field, & Pierce, 2015). This latter study included all effect sizes reported in correlation tables from primary studies and is therefore expected to find lower overall average effects as the prior studies looked only at hypothesized effects. Given these somewhat disparate findings, we formally state our first research question:

> **Research Question 1:** What is the average effect size obtained in micro-oriented management meta-analyses?

Cohen's Effect Size Categorization

Cohen (1962) first offered specific magnitudes for effects in an article published in the *Journal of Abnormal and Social Psychology*. In this article he described an effect size of .40 as medium because it was close to the average observed value found in his review of all articles published in that journal in 1960. Cohen went on to characterize an effect size of .20 as "small" and .60 as "large." Later, Cohen (1988) revised these values such that "small" effects are characterized by a $d = .2$ or $r = .1$, "medium" effects are characterized by a $d = .5$ or $r = .30$, and "large" effects are characterized by $d = .8$ or $r = .50$. However, in presenting this framework, Cohen (1988) cautioned:

> The terms "small," medium," and "large" are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation. In the face of this relativity, there is a certain risk inherent in offering conventional operational definitions for these terms for use in power analysis in as diverse a field of inquiry as behavioral science. This risk is nevertheless accepted in the belief that more is to be gained than lost by supplying a common conventional frame of reference which is recommended for use only when no better basis for estimating the ES index is available. (p. 25)

Despite this caution and the stated risk of making such general statements for a broad and diverse behavioral science literature, scholars have nearly ubiquitously adopted Cohen's effect size framework in interpreting the size of reported effects. It has become commonplace to describe one's reported effects as small, medium, or large (based on Cohen's categories) even though doing so ignores the relative nature of effect sizes as discussed above by Cohen himself. Glass, McGaw, and Smith (1981) emphasized this point:

> Above all else this is clear about magnitudes of effect: *There is no wisdom whatsoever is attempting to associate regions of the effect-size metric with descriptive adjectives such as "small," "moderate," "large," and the like.* [Italics in original] Dissociated from a context of decision and comparative value, there is no inherent value to an effect size. (p. 104)

Rosenthal (1994; Rosenthal & Rubin, 1979) has also expressed the dangers of comparing effect sizes using Cohen's framework and put forth alternative means for doing so (see Rosenthal & Rubin, 1982). Aguinis and his colleagues (Aguinis et al., 2010; Aguinis, Beaty, Boik, & Pierce, 2005) have likewise emphasized the subjective nature of Cohen's effect size categories. More recently, an empirical examination, based on effect sizes reported in *Journal of Applied Psychology* and *Personnel Psychology*, reported that Cohen's effect size categorization "bear almost no resemblance to findings in the field" (Bosco et al., 2015, p. 431).

There are, however, scholars who defend Cohen's effect size labels and encourage their use in various fields. For example, Aiken and West (1991) reference Cohen's work and repeat the conventional definitions he provides. Lipsey (1998) indicates that Cohen's rule of thumb for effect size magnitudes fits fairly well with Lipsey and Wilson's (1993) distribution of effect sizes in their review of 300 meta-analyses in psychology and education research. Cohen (1992), in reviewing his own effect size definitions, cites support for his "subjectively" (p. 156) derived effect size categories in various fields.

Much like NHST, despite much debate about Cohen's conventional definitions of effect size magnitude, they are used pervasively in many fields of study and organizational psychology is no exception (Aguinis et al., 2005). Given the uncertainty surrounding the appropriateness of Cohen's effect size categories, it is difficult to know their usefulness in OB/HR research.

> **Research Question 2:** Are Cohen's conventional definitions of effect size magnitude appropriate for use in interpreting effect sizes obtained in micro-oriented management research?

In order to answer this research question, we will look at the distribution of effect sizes in the OB/HR literatures from three different perspectives: The literature as a whole, divided by content areas, and relationships among content areas. The answer to our second research question will help micro-oriented management researchers decide whether or not the conventional practice of interpreting effect size magnitudes by referring to Cohen's work should be continued. Regardless, we are confident that our review of the meta-analyses in management research will provide an empirically derived means of making effect size comparisons specifically tailored to OB/HR research. Ferguson (2009) provided a similar critique of the appropriateness of Cohen's effect size magnitudes but his work is specific to clinical psychology. The ability to make meaningful comparisons among effect sizes is important because it allows researchers to move beyond statistical significance and toward practical significance (Fritz, Scherndl, & Kühberger, 2012). Moreover, because we benchmark the expected size of relationships between different categories of variables, we expand on prior research by providing specific estimates between particular independent and dependent variables rather than overall estimates of effect sizes across the field as a whole. This should be particularly useful for researchers conducting power analyses in order to estimate needed sample sizes to produce robust research findings.

Statistical Power

Just as management researchers have historically deemphasized effect sizes in favor of NHST, researchers have also been guilty of an imbalance in their treatment of the potential errors associated with NHST. A Type I error occurs when a true null hypothesis is rejected, with risk α, and a Type II error occurs when a false null hypothesis is sustained, with risk β. As previously discussed, OB/HR researchers routinely report their tests of significance and the associated α (typically .05). However, very infrequently is any mention made of the β (see Chase & Chase, 1976; Mazen et al., 1987; Mone, Mueller, & Mauland, 1996), or the risk of failing to reject a false null hypothesis. In other words, management researchers have given very little attention to the issue of statistical power (which is equal to $1 - \beta$).

Here again it is Cohen (1988, 1992) who has been the most outspoken advocate of increased attention to statistical power in the social sciences. Cohen's work suggested that the failure to detect a true effect (β) is about four times as serious as detecting an effect that is not true (α) and therefore a β of .80 was suggested in conjunction with the conventional α of .05 (Cohen, 1988). Cohen's work also provided a step-by-step procedure and the corresponding tables for calculating statistical power for numerous types of statistical tests. Cohen's initial statistical power analysis of the *Journal of Abnormal and Social Psychology* spurred similar efforts in management (Ferguson & Ketchen, 1999; Mazen et al., 1987; Mazen, Hemmasi, & Lewis, 2006; Mone et al., 1996), marketing (Sawyer & Ball, 1981), communications (Chase & Tucker, 1975; Kroll & Chase, 1975), psychology (Bakker, van Dijk, & Wicherts, 2012; Rossi, 1990; Sedlmeier & Gigerenzer, 1989), and education (Brewer, 1972; Penick & Brewer, 1972). These reviews unanimously found that social science research is underpowered and that Cohen's advocacy for explicit consideration of statistical power has had little effect on the actual practice of behavioral science researchers. Mazen et al. (1987) found a "general superiority" in the statistical power of management research when compared with psychology, education, and communications yet statistical power to detect small effects was only .31 (p. 372). In addition to being underpowered, management researchers rarely, if ever, make mention of statistical power (Cashen & Geiger, 2004; Mazen et al., 1987). Mone et al. (1996), noting these shortcomings, surveyed micro-oriented management and applied psychology researchers to assess their perceptions and usage of statistical power. They reported that 64% of the researchers surveyed never use statistical power analysis and the most common reason given for failing to do so is that editors and reviewers do not require it. Nearly half of those surveyed cited fixed sample sizes as another reason for not conducting statistical power analyses. Mone et al.'s (1996) findings help explain why statistical power is largely ignored in management research.

A number of years have passed since the most recent survey of statistical power in management and applied psychology was published (see Mone et al., 1996). More recently, Cashen and Geiger (2004) conducted an assessment of statistical power in management research, but their focus was specifically on the rather uncommon practice of hypothesizing a null relationship. In addition to the rather dated information available with regards to statistical power in management research, the prior reviews have been restricted to a limited number of sources and years. Additionally, prior reviews have focused exclusively on Cohen's small, medium, and large effect sizes when calculating power with little or no mention of the actual effect sizes obtained in management research. An exception to this is Mazen et al. (1987) who used the average effect size ($d = .39$) obtained in 12 meta-analyses as a proxy for the actual effect sizes obtained in the 84 studies they reviewed. Our review expands on and improves this methodology by including a more comprehensive set of meta-analyses and using actual effect sizes rather than the subjective categories set forth by Cohen. The more comprehensive nature of our data will allow us to make a more thorough and up-to-date assessment of the statistical power in OB/HR research. Thus, we will be able to answer our next research question:

> **Research Question 3:** What is the statistical power of the average study in micro-oriented management meta-analyses?

Statistical power is an important consideration because the consequences of ignoring it are potentially damaging to the advancement of a field of study. One the one hand, ignoring statistical power can result in seemingly conflicting findings (Howard, Maxwell, & Fleming, 2000). Underpowered studies often fail to detect findings that well-powered studies detect. This occurrence hinders the accumulation of scientific knowledge. One way that it does this is by bringing into doubt any evidence that appears to falsify a theory, which is an important means of scientific advancement (Greenwald, 1993). That is, anomalous research findings can simply be attributed to the high sampling error that is associated with small sample sizes. On the other hand, ignoring statistical power can result in sample sizes that far exceed what is required to detect effects. This is a waste of time and resources that could otherwise be allocated to additional research questions (Mazen et al., 1987). Moreover, larger-than-necessary samples can render nearly any effect statistically significant regardless of how trivial in nature it is. In sum, despite being largely ignored, issues of statistical power have real and lasting consequences.

The File Drawer Problem

Another problem that results from the primary focus in the social sciences on statistical significance is publication bias. Broadly defined, publication bias suggests that the pool of studies deemed worthy of publication is not representative of all research that is conducted. Sutton (2009) describes several potential sources of publication bias and all of them, in one way or another, relate to the statistical significance of findings. As a result, publication bias typically refers to the tendency for statistically significant findings to be published over nonsignificant findings (Rosenthal & Rubin, 1979). In other words, because of the bias against nonsignificant findings, a sample of published studies inaccurately represents the research population. The general tendency of a publication bias, then, is to over-state the magnitude of effects (as larger effect sizes will be statistically significant, all else being equal) or report effects that do not actually exist. This effect is further strengthened by underpowered studies because the effect size that is needed to observe a statistically significant effect with an underpowered study is often far greater than the actual population effect size (Cohen, 1994).

Because publication bias occurs at the primary study level, any attempt to synthesize or review primary studies will also be subject to this bias to the extent to which they draw on published research. This is often referred to as the file drawer problem (Greenwald, 1975; Rosenthal & Rubin, 1979) because studies with nonsignificant findings are often relegated to a file drawer and neither submitted for publication nor offered up to meta-analysts for inclusion in reviews. That publication bias exists and has some effect on meta-analyses is widely accepted (Rothstein, Sutton, & Borenstein, 2005; Sutton, 2009), but the magnitude of this effect is contested (Hunter & Schmidt, 2004).

Publication bias in meta-analysis has been described as a "major threat" (Sutton, 2009, p. 436), an "800-lb gorilla" (Ferguson & Heene, 2012, p. 556), and "widespread and pervasive" (Levine, Asada, & Carpenter, 2009, p. 290). For example, Ferguson and Brannick (2012) found that approximately 25% of meta-analyses were at risk of publication bias (Ferguson & Heene, 2012). But, based on a negative correlation between effect size and sample size found in 80% of meta-analyses they reviewed, Levine et al. (2009) suggested that the incidence of publication bias is much higher.

Recently, Dalton, Aguinis, Dalton, Bosco, and Pierce (2012) offered a different view regarding the incidence of publication bias in meta-analyses. These authors suggest that publication bias may not be as great a threat to meta-analysis as commonly believed, basing this claim on the fact that most empirical analyses in contemporary research is multivariate in nature and includes a correlation table that reports relationships not just among the hypothesized variables but all other variables as well. This table, then, often includes nonsignificant results that would be immune from the hypothesized publication bias because they were not relevant to the research questions of the primary study. Dalton et al. (2012) cite several examples that illustrate that these correlation tables are likely to be included in meta-analyses even though the relationship of interest to the meta-analyst was not of interest to the author of the primary study. Thus, meta-analyses are likely to include both significant and nonsignificant results even if they only include published research. Indeed, a comparison of published versus unpublished correlation tables allowed Dalton et al. (2012) to conclude,

> Given that there are no differences regarding the pre-specified Type I error rates, sample size, and percentage of nonsignificant correlations across data sets, we can conclude that there are no differences in the magnitude of correlation coefficients comparing published versus nonpublished effect sizes. (p. 241)

More recently, Stemig and Sackett (2013) and Kepes, Banks, McDaniel, and Whetzel (2012) largely confirmed the findings of Dalton et al. (2012) but stopped short of labeling publication bias a nonissue.

Clearly disagreement remains about the size of the threat that publication bias poses to meta-analysis. However, Dalton et al.'s (2012), Kepes et al. (2012) and Stemig and Sackett's (2013) conclusions that publication bias does not present a serious threat are more recent, robust and supported to a greater degree by empirical findings than are those that claim otherwise. Thus, we formally hypothesize:

Hypothesis 1: Inclusion of unpublished studies will have no effect on the effect size reported in micro-oriented management meta-analyses.

This hypothesis is important because "the file drawer problem and related issues are easily among the most extensively chronicled critiques of the responsible interpretation of meta-analyses, or lack thereof" (Dalton et al., 2012, p. 226). In fact, so widespread and accepted has been the threat of publication bias in meta-analysis that numerous and varied approaches have been devised to correct for it (see Sutton, 2009, for a review). In addition, there has been a heated debate recently regarding the effect that including unpublished studies in meta-analyses has on publication bias (see Ferguson & Brannick, 2012; Rothstein & Bushman, 2012). Ferguson and Brannick (2012) claim that inclusion of unpublished studies actually increases the bias in meta-analyses whereas Rothstein and Bushman (2012) claim that this finding is due to idiosyncratic choices and led to an erroneous conclusion.

Methodology

Document Retrieval

Articles for inclusion in our review were identified by examining the contents of journals that appeared on the list of Eigenfactor.org's top 30 most impactful management journals and had published at least one meta-analysis by June 1, 2012. We then used Academic Search Premier to search these journals for articles whose titles included the terms *meta-analysis, meta-analyses, meta-analytic*, or *quantitative review*. This search yielded 350 articles appearing in eleven journals: *Academy of Management Journal*, *Human Performance*, *Journal of Applied Psychology*, *Journal of Management*, *Journal of Management Studies*, *Journal of Occupational and Organizational Psychology*, *Journal of Organizational Behavior*, *Organizational Behavior and Human Decision Processes*, *Organization Studies, Personnel Psychology*, *and The Leadership Quarterly*.

Criteria for Inclusion

Articles were included in our review if they satisfied three criteria: (1) the article reported a meta-analytic estimate of the strength of the relationship between two variables, (2) the estimate was based on at least three independent samples (i.e., $k > 2$) from at least two sets of authors, and (3) the reported meta-analytic effect size fell into the "micro" domains of OB, HR, or Industrial and Organizational Psychology (I-O). Some of the journals we included in our initial document reported meta-analytic findings for constructs outside the realm of OB/HR/I-O. We therefore eliminated studies on the basis of content. The first author reviewed each meta-analysis that was part of our initial pool and made inclusion decisions in consultation with the other authors. This procedure resulted in the elimination of publications that were clearly focused on strategic management, business policy, or related topics. Using these procedures, we eliminated 92 articles resulting in a final list of 258. The majority of meta-analyses eliminated clearly belonged to more "macro" disciplines (e.g., business policy and strategic management). For example, we did not include Dalton, Daily, Johnson, and Ellstrand's (1999) meta-analysis on the number of directors and financial performance as the variables of interest were clearly at the firm level. Other studies eliminated from our database included the search words in the title but were not actually meta-analyses. For example, Aguinis et al. (2011) was eliminated for this reason.

Coding of Effects

We sought and coded the most highly aggregated effect from each meta-analysis. When a meta-analysis did not provide a single, summary effect, we coded up to six of the most highly aggregated effects from each meta-analysis. For example, in meta-analyses where Big 5 personality traits were a variable of interest, we coded correlational relationships for all five personality traits rather than just capturing a single relationship. When more than six highly aggregated effects were present, we selected effects that had been studied with higher relative frequency and which had not yet been captured in our coding of previous meta-analyses. The decision to include more than one effect from some meta-analyses introduces questions regarding independence of effects. However, we followed the methodological recommendation of Aguinis et al. (2005; Aguinis et al. 2011) and conducted our analyses at the effect-size level instead of the meta-analytic article level.

The statistics we coded from each meta-analysis are the mean uncorrected correlation (r), the population correlation (i.e., ρ), the observed variation in effect sizes (i.e., SDr), and the observed variation in effect sizes after accounting for artifacts (i.e., $SD\rho$). For studies that did not report the variance directly, we coded either the credibility interval or the Q statistic. From each study we also coded the number of independent samples (k), the overall sample size (N) and the proportion of studies that had not been published. Finally, to aid in the interpretation of the distribution of ρ and $SD\rho$ we also coded the artifacts that each meta-analysis corrected for (e.g., unreliability in measurement of independent and/or dependent variable, range restriction, interrater reliability, artificial dichotomization of variables).

The coding and double-coding was completed by the first author. Following the first round of coding, the first author recoded the entire sample of meta-analyses to ensure the accuracy of the coded values used in our analyses (r, ρ, SDr, $SD\rho$, k, N). To further ensure the accuracy of coding, a random sample of 21 studies (or approximately 8%) was also coded by the fourth author. The resulting interrater reliability was 97%. Discrepancies in coding were resolved by referring to the original meta-analysis in question and agreement was reached by the authors regarding the correct value to code.

Analyses

When a meta-analysis in our database reported the effect size in another form, we converted the effect size to *r.* We also converted variance estimates from meta-analyses expressed in *d* effect sizes to that of correlations. There is some controversy regarding the ability to convert variance estimates from one metric to another. In particular, Schmidt (2008) states,

> Although any value of Fisher's *z* or any mean Fisher's *z* can be converted to *r*, it is not possible to convert a variance or SD in Fisher's *z* metric to the *r* metric. This is a serious limitation for all methods using the Fisher's *z* metric, because to be usable the final results have to be expressed in the correlation metric. (p. 111)

Though this is long established position held by several others (e.g., S. M. Hall & Brannick, 2002; Schulze, 2004), this is no longer the case. As Steel (2013) documents, conversion of variance scores among those based on Fisher's *z*, correlation or *d* scores can be readily done through a mathematical process called *numerical integration*. Applied by Steel and Kammeyer-Mueller (2008), who used it to calculate the probability density for Bayesian meta-analytic variance estimation, essentially the range of the original distribution is divided into hundreds of small strips or slices, each associated with a probability (i.e., calculated by the probability density function). When these strips are made small enough that they can be summarized by a single point (i.e., more precise figures slip below rounding error), these points are converted into the target metric (e.g., *z* to *r* or *r* to *d*). Then, each of these newly converted points is multiplied by their associated strip probability and sum, arriving at the mean average. The variance is simply the sum of the squares around this mean for each strip, multiplied by the probability of each respective strip.

Many variance estimates were also reported in the form of credibility intervals or *Q* statistics. The credibility intervals were converted to an estimate of variance. The *Q* statistics were also converted to variance estimates by multiplying the *Q* statistic by sampling error and dividing by *k*.

Results

The 258 meta-analyses that were included in this review yielded 776 meta-analytic conclusions. Based on these meta-analytic conclusions and their associated metrics, our hypotheses and research questions were tested as described below.

Magnitude of Effect Sizes

Research Question 1 focused on the average effect size in micro-oriented management meta-analyses. To answer this research question, we will report both uncorrected and corrected effect sizes. First, based on 686 uncorrected effect sizes we calculated an average effect size (r) of .227 (SD = .145, median r = .200). This finding is somewhat higher than that reported by Mazen et al. (1987; r = .19) and Bosco et al. (2015; r = .16), but as explained earlier, the latter study averaged all reported effect sizes including those not hypothesized. Next, using the 690 corrected effects (ρ) included in our data, we calculated an overall mean effect size value of ρ = .278 (SD = .174, median ρ = .256). This value is similar to the overall mean effect size reported by Aguinis et al. (2011; r_c = .261) and the somewhat higher value we report can be at least partially explained by our exclusion of macro-level variables, which tend to be associated with lower effect sizes.

In addition to calculating the overall mean effect size, we also calculated the average mean effect size by topic area. These findings are presented in Table 1. To compile this table we assigned each variable involved in a meta-analytic conclusion a topic area. Therefore, each effect size is represented twice in Table 1 except where both variables in a meta-analytic conclusion belong to the same topical category. For example, a meta-analysis that reported a relationship between a job attitude and performance would be included in the average effect size calculation for the job attitude category and the performance category. As Table 1 illustrates, demographic variables (e.g., gender, age, education level) were found to have the smallest relationships with other variables (ρ = .115). This finding is consistent with Richard et al.'s (2003) review of meta-analyses that found the effect size of gender differences were relatively small in social psychological research as well (ρ = .12).Leadership constructs and culture, climate, and structure variables were found to have the strongest relationships with other variables (ρ = .35).

Next, we take a more fine-grained look at the effect sizes within each topical category. Table 2 displays the average corrected and uncorrected correlations found between each of the 19 topical categories of variables as well as the corrected standard deviation of effect sizes. In addition, the table displays the total k and N for each relationship along with the recommended minimum sample size for statistical power at .80 and .95 based on the overall uncorrected effect size. An inspection of the table reveals that leadership variables had among the largest effect sizes with three major dependent variables: job attitudes (.56), performance (.27), and turnover (.47). This finding potentially provides evidence for the predictive strength of leadership theories and measures. An alternative interpretation, and one that is more skeptical of leadership constructs, is that leadership's high correlations with other variables is driven mainly by the fundamental attribution error. In other words, when workers are satisfied with their jobs they tend to overattribute this to their leaders (see Bligh, Kohles, Pearce, Justin & Stovall, 2007; Meindl, Ehrlich & Dukerich, 1995; Weber, Camerer, Rottenstreich, & Knez, 2001).

Table 1. Effect Sizes, Variance, and Statistical Power by Meta-Analytic Article Topic.

Topic	No. of conclusions	Total *k*	$\lvert\bar{r}\rvert$	*SDr*	$\lvert\bar{\rho}\rvert$	$SD\bar{\rho}$	SP	SP w/o outliers
Attitudes	186	6,606	.29	.15	.34	.13	.84	.72
Culture, climate, structure	36	1,503	.23	.14	.35	.16	.69	.60
Creativity, innovation, learning	27	693	.14	.12	.17	.20	.39	.39
Demographic variables	86	4,068	.12	.09	.15	.11	.45	.42
Deviant behaviors	16	835	.24	.10	.30	.01	.88	.76
Extra-role behaviors	40	1,274	.21	.14	.25	.12	.71	.67
Human resource practices	87	6,220	.23	.11	.31	.15	.68	.61
Individual differences	216	8,893	.19	.13	.24	.13	.61	.56
Interpersonal processes	47	1,470	.23	.15	.22	.13	.74	.65
Job characteristics and design	73	1,532	.24	.10	.29	.10	.79	.69
Leadership	69	2,261	.29	.11	.35	.16	.78	.69
Motivation	31	1,112	.21	.12	.25	.12	.60	.56
Performance evaluation	197	10,581	.18	.14	.24	.13	.55	.51
Perceptions	70	2,836	.27	.12	.32	.13	.80	.64
Stress and aggression	36	1,518	.24	.14	.29	.15	.80	.68
Safety and health	28	627	.21	.12	.26	.15	.82	.73
Training	32	1,314	.19	.13	.25	.21	.52	.50
Turnover	61	1,831	.21	.12	.24	.11	.74	.66
Teams and groups	61	1,646	.21	.13	.27	.15	.47	.42

Note. SD = standard deviation; SP = statistical power; *k* = number of studies; $\lvert\bar{r}\rvert$ = average absolute value uncorrected effect size; *SDr* = standard deviation of reported average uncorrected effect sizes; $\lvert\bar{\rho}\rvert$ = average absolute value corrected effect size; $SD\bar{\rho}$ = average standard deviation of corrected effect sizes.

A general pattern evident in the table is that variables that are usually other-rated or objective (i.e., turnover and performance) tend to have smaller relationships with other variables (which are typically self-report). This finding, while not surprising, is additional evidence of the same-source bias that plagues much of OB/HR research. Because of the sheer size of Table 2, an interpretation of each cell is not reasonable, however, each individual cell could be of importance to researchers interested in heeding Edwards' (2008) advice to formulate hypotheses that go beyond directionality by specifying a range of potential values for the predicted relationship. Finally, the blank cells in Table 2 indicate relationships between two types of variables that have not yet been meta-analyzed (e.g., leadership and creativity), or they have not been published in the journals that we included in our review. These blank cells may also indicate areas where there are a paucity of primary studies and where scholarly attention should be directed.

Cohen's Effect Size Categorization

Research Question 2 dealt with Cohen's system for categorizing effect sizes. According to Cohen (1988), "small" effects are characterized by a $d = .2$ or $r = .1$, "medium" effects are characterized by a $d = .5$ or $r = .30$, and "large" effects are characterized by $d = .8$ or $r = .50$. To investigate whether or not our data support this classification system, we created a histogram that depicts the frequency with which various uncorrected effect sizes were reported (see Figure 1). Figure 2 provides similar information for corrected effects, but in this section we interpret only the uncorrected effects as this will be most useful to researchers desiring to compare their reported (uncorrected) effect sizes to others. In Table 3, we provide additional detail regarding the breakdown of effect sizes by percentile. Cohen's "small" effect size ($r = .10$) appears at the 20th percentile, indicating that an effect size of .10 is smaller than 80% of effect sizes reported. Cohen's "medium" effect size ($r = .30$) appears near the 75th percentile, indicating that an effect size of .30 is larger than all but 25% of reported effect sizes. Cohen's "large" effect size ($r = .50$) appears near the 95th percentile, indicating that an effect size of .50 is larger than 95% of reported effect sizes. Based upon this analysis, it appears that Cohen's categorization is not ideal for describing effect sizes in micro-oriented management research because the vast majority of effect sizes fall between his categories labeled "small" and "medium." Our finding here is similar to that reported by Bosco et al. (2015) in that Cohen's benchmarks do not accurately reflect the actual distribution of effect sizes in micro-oriented management research. The percentile distribution reported in Table 3 can be used by researchers in reporting effect sizes in lieu of Cohen's labels. For example, a researcher reporting an effect size of .15 could report that this effect size exceeds in magnitude approximately 35% of the effect sizes reported in the HR and OB literatures. This reporting provides greater specificity than the common practice of describing effect

Table 2. Correlations Among Content Areas.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Attitude (*r*, *k*)	.40, 407																		
ρ, *SD*ρ	.50, .16																		
N	170,071																		
SS (β = .20, .05)	37, 63																		
2. Culture (*r*, *k*)	.30, 190	—, 124																	
ρ, *SD*ρ	.36, .15	.48, .19																	
N	74,887	9,201																	
SS (β = .20, .05)	67, 116	—, —																	
3. Creativity (*r*, *k*)	—	22, 88	—																
ρ, *SD*ρ		—,—																	
N		4,729																	
SS (β = .20, .05)		126, 219																	
4. Demo. (*r*, *k*)	.08, 774	.04, 19	.13, 23	—															
ρ, *SD*ρ	.10,.13	.04, .10	.14, .28																
N	7,08,708	4,073	8,877																
SS (β = .20, .05)	964, 1k	1k, 1k	3,64, 636																
5. Deviance (*r*, *k*)	.25, 85	33, 21	—	.15, 66	.52, 27														
ρ, *SD*ρ	.29, .11	36, .09		—, —	.62, .11														
N	5,7270	50,509		33,164	10, 104														
SS (β = .20, .05)	97, 168	55, 95		273, 476	21, 35														
6. Xtra-role (*r*, *k*)	.20,371	30, 18	—	—	.27, 49	—													
ρ, *SD*ρ	.24, .07	36, .12			.32, .34														
N	66,976	7,490			16,721														
SS (β = .20, .05)	1,53, 266	67, 116			83, 144														
7. HR (*r*, *k*)	.24, 482	—	—	.17, 306	.33, 443	—	28, 1,639												
ρ, *SD*ρ	.40, .16			.17, .12	.47, .37		32, .16												
N	134,598			8,594	507,688		290,044												
SS (β = .20, .05)	106, 183			212, 370	55, 95		77, 133												
8. Indiv. D. (*r*, *k*)	.21, 953	26, 13	.10, 1299	.09, 317	.16, 66	.15, 278	20, 1,809	.23, 873											
ρ, *SD*ρ	.25, .13	32, .10	.11, .13	.06, .09	.19, .09	.17, .11	30, .17	.27, .16											
N	275,152	1,891	28,686	497,288	13,625	56,027	212,299	219,272											
SS (β = .20, .05)	139, 241	90, 155	617, 1k	762, 1k	240, 418	273, 476	153, 266	115, 200											
9. In, -Person (*r*, *k*)	27,318	23, 27	—	—, 246	—	.27, 8	—	.27, 37	.11, 227										
ρ, *SD*ρ	.36, .10	30, 27		.04, .17		33, .21		.31, .12	.18, .14										
N	93,654	3,554		61,496		2,011		13,779	38,638										
SS (β = .20, .05)	83,144	1,15,200		—, —		83, 144		83, 144	509, 890										

(continued)

Table 2. (continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
10. Job Des. (*r*, *k*)	.35, 444	—	—	−, 75	—	.17, 57	—	.25, 115	.19, 248	—, 107									
ρ, *SD*ρ	.44, .08			.09, .12		.21, .13		.33, .09	.20, .07	.27, .20									
N	204,805			80,363		14,174		39,420	107,440	55,375									
SS (β = .20, .05)	49, 84			—, —		2,12, 370		97, 168	170, 295	—, —									
11. Leader (*r*, *k*)	.45, 400	—	—	—	—	.25, 103	—	.18, 589	.33, 34	—	.33, 785								
ρ, *SD*ρ	.56, .09					.30, .10		.23, .13	.38, .21		.39, .21								
N	90,946					17,596		125,462	6,683		166,492								
SS (β = .20, .05)	29, 49					97, 168		189, 329	55, 95		55, 95								
12. Motiv. (*r*, *k*)	—	—	—	.06, 216	—	—	—	.20, 50	—	.32, 12	—	.23, 62							
ρ, *SD*ρ				.08, .11				.24, .07		.52, .06		.30, .13							
N				181,197				6,124		10,203		3,765							
SS (β = .20, .05)				1k, 1k				153, 266		59, 101		115, 200							
13. Percep. (*r*, *k*)	.35, 407	—	—	—	—	.15, 79	42, 101	.19, 20	.25, 149	—	—	—	.38, 766						
ρ, *SD*ρ	.42, .12					.18, .04	.49, .14	.24, .08	—, —				.42, .20						
N	161,544					18,463	31,231	3,313	81,879				183,926						
SS (β = .20, .05)	49, 84					273, 476	33, 57	170, 295	97, 168				41, 70						
14. Perfor. (*r*, *k*)	.14, 601	13, 132	—	.11, 1401	—	.62, 184	25, 1258	.17, 2825	.06, 74	.16, 220	.23, 177	.25, 690	.15, 668	.28, 636					
ρ, *SD*ρ	.19, .08	24, .12		.14, .14		.67, .26	26, .15	.24, .13	.05, .07	.18, .10	.27, .18	.26, .11	.18, .12	.38, .17					
N	1,17,726	23,451		1.475M		63,531	261,182	474,589	24,060	62,532	21,120	48,366	12,1978	98,378					
SS (β = .20, .05)	97, 168	364, 636		509, 890		15, 23	97, 168	212, 370	1k, 1k	240, 418	115, 200	97, 168	273, 476	77, 133					
15. Stress (*r*, *k*)	.25, 243	—	.04, 82	—, 19	.22, 25	.05, 19	—	.32, 62	.46, 63	.11, 68	—	—	.27, 472	.12, 93	.37, 222				
ρ, *SD*ρ	.35, .17		.04, .19	.01, .11	.28, .06	.06, .18		.38, —	.56, —	.11, .09			.33, .12	.16, —	.46, .19				
N	74,858		7,417	11,364	7,343	6,022		15,860	195,78	33,538			100,209	18,408	47,516				
SS (β = .20, .05)	97,168		1k, 1k	—, —	1,26,219	1k, 1k		59, 101	28, 46	509, 890			83, 144	428, 747	44, 74				
16. Safety (*r*, *k*)	.13, 9	36, 64	—	—	.18, 29	—	—	.19, 231	—	.13, 107	.12,28	—	.11, 40	—	—	.35, 119			
ρ, *SD*ρ	.17, .05	.45, .11			.27, .07			.28, .22		.15, .13	.14, .19		.12, .08			.42, .17			
N	20,078	47,419			45,880			54,444		149,708	17,849		53,530			45,316			
SS (β = .20, .05)	364, 636	46, 79			189, 329			170, 295		364, 636	428, 747		509, 890			49, 84			
17. Train. (*r*, *k*)	.20, 104	.05, 36	.16, 221	.08, 16	—	—	—	.21, 39	—	.30, 35	—	.40, 65	—	.17, 697	—	—	—, 101		
ρ, *SD*ρ	.24, .17	.06, .19	.30, .28	.09, .13				.24, .13		.36, .19		.49, .22		.24, .17			.18, .25		
N	8,966	1,202	25,019	1,887				5,503		5,017		10,203		57,074			7,207		
SS (β = .20, .05)	153, 266	1k, 1k	240, 418	964, 1k				139, 241		67, 116		37, 63		212, 370			189, 329		
18. Turno. (*r*, *k*)	.32, 637	—	—	.07, 81	.29, 24	.17, 90	.10, 168	.16, 197	—	.17, 5	.40, 17	—	.19, 134	.18, 96	.20, 150	—	—	.24, 232	
ρ, *SD*ρ	.28, .12			.07, .10	.35, .11	.22, .17	.16, .06	.18, .06		—, —	.47, .12		.23, .10	.22, .17	.24, .14			.30, .13	
N	157,536			54,333	13,961	26,510	62,897	45,072		1,242	3,297		59,378	32,951	49,506			51,034	
SS (β = .20, .05)	59, 101			1k, 1k	72, 124	212, 370	617, 1k	240, 418		212, 370	37, 63		170, 295	189, 329	153, 266			106, 183	
19. Team (*r*, *k*)	.38, 144	—	.16, 75	.05, 335	—	—	—	.11, 213	.44, 39	.16, 34	—, 110	.31, 17	—	.22, 616	—	—	—	—	.24, 146
ρ, *SD*ρ	.43, .22		.18, .24	.21, .07				.15, .13	.52, .22	.21, .17	.30, .20	.37, .26		.26, .15					.27, .23
N	16,587		16,615	52,700				12,788	2,999	2,737	6,461	860		72,132					7,226
SS (β = .20, .05)	41, 70		240, 418	1k, 1k				509, 890	30, 51	240, 418	—, —	63, 108		126, 219					106, 183

Note. r = average uncorrected effect size; *k* = total number of effect sizes; ρ = average corrected effect size; *SD*ρ = average standard deviation of corrected effect sizes; *N* = total sample size; SS = minimum recommended sample size, reported for both β = .20 (or statistical power = .80) and β = .05 (or statistical power = .95); 1*k* = recommended sample size is greater than 1,000.

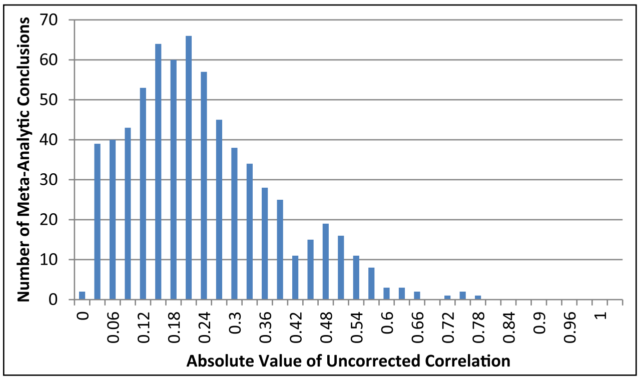

Figure 1. Frequency distribution of *r*.
Note. Mean = .227; median = .200; *SD* = .144; maximum = .770; minimum = .000; count = 686.

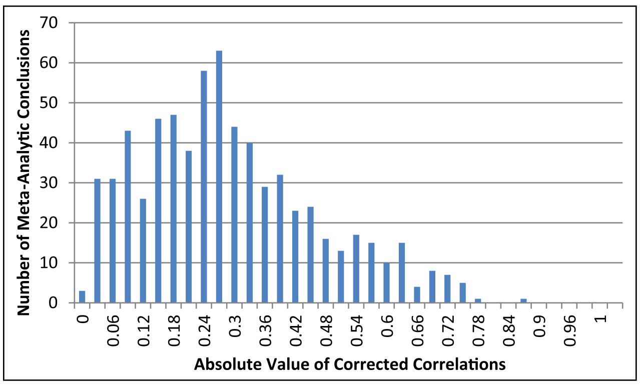

Figure 2. Frequency distribution of ρ.
Note. Mean = .278; median = .251; *SD* = .174; maximum = .850; minimum = .000; count = 690.

sizes as either small, medium, or large. Moreover, the information about the average effect size obtained by topical category (Table 1) and between two variables of different topical categories (Table 2) provide the ability to make even more fine-grained analysis of effect sizes. For example, a researcher reporting an effect size of .35 between job attitude and performance variables could report that this effect size is substantially larger than the average effect size reported between variables of these types (r = .14, ρ = .19) in prior studies. This type of reporting allows researchers to move beyond merely the statistical significance of a finding and gets closer to describing the practical significance of a finding when compared to other similar findings previously reported.

Statistical Power

Research Question 3 was concerned with statistical power. Given the concerns that have been raised about whether studies in the social sciences have sufficient sample sizes in order to detect their average effects (see Cohen 1962, 1990, 1992) we report the average statistical power of the studies contained in the meta-analyses summarized in this report (see Table 1). This information was derived by taking the sample size and the effect size from each meta-analytic conclusion and then calculating the statistical power based on β = .80. The average power of the 670 uncorrected effect sizes was .684. However, we should point out that due to inclusion of studies with extremely high sample sizes in numerous meta-analyses, we found that 156 of the 670 meta-analytic conclusions reported perfect statistical power (1.0). Therefore, we also calculated the average statistical power after removing these meta-analytic conclusions. This calculation resulted in a significantly lower estimate of the average statistical power (.588).

In his seminal paper on the subject of statistical power, Cohen (1992) suggests that power of .80 was necessary to avoid too great a risk of a Type II error (the failure to reject a false null hypothesis). Only 51.6% of the meta-analyses in the present review reported average statistical power levels above this convention. Put another way, roughly half of meta-analyses published in the micro-oriented management literature are based on studies that are, on average, statistically underpowered. We also calculated this percentage after removing "outlier" meta-analytic conclusions reporting perfect power (assuming that this extremely high statistical power may have been due to inclusion of very large sample sizes in the meta-analysis). This analysis revealed an even more worrisome picture of statistical power in OB/HR studies as only 37.0% of studies were found to meet the .80 mark.

Table 3. Distribution of Uncorrected Effect Sizes.

Percentile	Effect size
5	.03
10	.06
15	.08
20	.10
25	.12
30	.14
35	.15
40	.17
45	.19
50	**.20**
55	.22
60	.24
65	.26
70	.28
75	**.31**
80	.34
85	.37
90	.45
95	.51
100	.77

Note. Boldface numbers represent quartile partitions.

Table 4. Correlations Among Coded Variables.

Variable	1	2	3	4	5	6
1. *k*	*r*(*N*)					
2. *N*	.10* (762)	—				
3. ρ	.01 (689)	−.03 (688)	—			
4. *SD*ρ	.11* (560)	.10* (559)	.26** (537)	—		
5. Year of publication	−.05 (776)	−.01 (762)	−.08* (690)	.07 (562)	—	
6. %Unpublished	.02 (376)	.14** (366)	−.07 (330)	.04 (287)	.25** (456)	—

*$p < .05$. **$p < .01$.

The final two columns in Table 1 present information about average statistical power by topic area. As can be seen, topic areas that typically report smaller effect sizes, as expected, exhibit lower average statistical power (see creativity, learning, and innovation, for an example). Finally, Table 2 includes some additional information relating to statistical power. Each entry in the correlation table includes recommended minimum sample size recommendations for researchers based on the average uncorrected effect size for relationships between variables from two topic areas. These minimum recommended sample sizes are reported for statistical power equal to.80 and .95.

The File Drawer Problem

Hypothesis 1 stated that there would be no relationship between the percentage of unpublished studies included in a meta-analysis and the resulting effect size reported. We tested this hypothesis in two ways. First, we analyzed the degree to which inclusion of unpublished studies in meta-analyses influences the size of the meta-analytic effect reported. For relationships that reported the percentage of unpublished studies on which the estimate was based ($N = 456$) we calculated a correlation coefficient with the reported corrected ρ. As reported in Table 4, there was no significant relationship ($r = -.065$, $p > .05$) between the percentage of unpublished articles and the effect size reported. Second, we looked at the relationship between sample size and effect size in our sample of meta-analyses. In prior reviews, some have found a negative relationship between sample size and effect size in meta-analyses (e.g., Levine et al., 2009) and used this to suggest that publication bias is a significant problem in meta-analyses. As reported in Table 4, we did not find a statistically significant relationship between ρ and N in our study ($r = -.026$, $p > .05$). Thus, it appears that publication bias is not a major threat in the micro-oriented management literature. These findings are consistent with Hunter and Schmidt's (2004) argument that the file-drawer problem is likely to be of trivial size and confirm Dalton et al.'s (2012) findings that "the file drawer problem does not produce an inflation bias and does not pose a serious threat to the validity of meta-analytically derived conclusions" (p. 222).

Discussion

As the management literature matures as a field of inquiry, meta-analytic methods will become ever more important tools for researchers attempting to synthesize and bring clarity to research on a particular relationship. This article has aimed to summarize some of the key findings of the past 30 years of meta-analytic research in the OB/HR literatures, highlight differences among areas of inquiry, and examine research tendencies in the general domain. Specifically, we were interested in the magnitude and variability of the effects micro-oriented management researchers study, how these effects vary by area of inquiry, the file drawer problem, and the average statistical power reported in meta-analytic conclusions.

At the broadest level the typical meta-analysis published in the past 30 years is one based on a synthesis of a relatively large literature (mean $k = 43.5$, mean $N = 32{,}731$) reporting an average uncorrected effect size of .227 and a corrected population correlation of .278. The average effects reported here are similar to the .261 mean corrected ρ in Aguinis et al.'s (2011) review of 196 meta-analyses in management research and Richard et al.'s (2003) finding an average r of .21 in the social psychology literature. Importantly, the present study goes beyond estimating the *overall* expected effect size for management (e.g., Aguinis et al., 2011) and reports the expected effect size of relationships *between* particular topics in management at a more fine-grained level than previous attempts (e.g., Bosco et al., 2015). This is important when considering what the expected effect size for research and practice is because the present results illustrate substantial variability in expected effect sizes *across* topic areas. For example, stress and aggression research typically has an effect size of .29 while demographics-focused research has an effect size of .17. But even *within* topic areas, we find substantial variability as well. Once again, consider the case of stress and aggression research. The expected effect size with positive extra-role behaviors is .06, but the expected effect size with workplace deviance is .28. Clearly, there is a need to consider the particular relationships under investigation. Although Table 2 does not provide estimates for every possible relationship, the present results should provide needed guidance for future research not only in providing more accurate benchmarks for expected effect sizes, but also by illustrating where further meta-analytic research is most needed.

In addition to reporting the average magnitude of effect sizes in the micro-oriented management literature, we also provide an analysis of the distribution of effect sizes. Based on this information we assessed the suitability of the commonly-used effect size categories put forth by Cohen (1988). We conclude that his effect size categorization system is largely inappropriate for the OB/HR literatures because it is not based on the actual distribution of effect sizes reported in this literature. We suggest that researchers refer to the effect size distributions reported in this study in order to make more informed and meaningful comparison between effect sizes in the management literature.

It is also important to note that many of the examined meta-analyses synthesize individual studies that are often characterized by very low power, thereby greatly increasing the value of these meta-analyses since the Type 2 errors associated with underpowered studies can be avoided when synthesizing numerous such studies. The low power of the average study synthesized by some of the examined meta-analyses (average sample size, average effect) is obviously of some concern given the long standing exhortations (e.g., Asendorpf et al., 2013; Mazen et al., 1987; Mone et al., 1996) to design studies with adequate power, but must be weighed against the practical difficulties of obtaining large samples in many domains of management research (e.g., leadership research, group climate research). Obviously this is an issue that confronts many social science fields. Our finding that the statistical power for the average study represented in the typical meta-analytic conclusion is .588, after removing outlying meta-analyses reporting perfect power, is largely consistent with the .50 reported by Maxwell (2003) for the field of psychology, though larger than the .35 reported by Bakker et al. (2012). As Adendorpf et al. (2013) report, this bonanza of low powered studies is leading to excessive number of false-positive findings, damaging our field as a whole. Use of Table 2, which provides recommended sample sizes for adequate power, by researchers, editors, and reviewers should address this.

Our study also contributes to the conversation regarding the file drawer problem. The file drawer problem suggests that effect size estimates are often overstated due to the publication bias toward statistically significant results. Our analysis concludes that the inclusion or exclusion of unpublished studies does not significantly affect the effect size reported in meta-analyses. This finding confirms Hunter and Schmidt's (2004) opinion that the file drawer problem is trivial, at least within the OB/HR fields, and is consistent with a recent study by Dalton et al. (2012) that reports a similar conclusion. For other areas of inquiry, especially medicine (e.g., R. Hall, Antueno & Webber, 2007; Song et al., 2009), publication bias can be of greater concern.

Limitations

Although we included a fairly broad literature and a considerable number of studies, our project is not without its limitations. By limiting our analysis to subjects that have been meta-analyzed and published in the top journals, we may have left unaddressed various topics that are of interest to management researchers. Moreover, a number of meta-analyses published in lower tier journals that may have been relevant for the present analysis were excluded. One potential criticism of this decision is that the present results are reflective only of meta-analyses published in premier management journals. However, it could also be argued that this decision rule meant that the present results are more likely to be based on high-quality meta-analyses of topics deemed to be of serious import to management scholars. Moreover, because meta-analyses typically aggregate across many potential sources of data, the present results are still likely to be reflective of research findings in management as a whole.

A related concern is that we have chosen to include only a limited number of meta-analytic conclusions from each meta-analysis in our study. Although this decision resulted in fewer conclusions being included in our final analysis, the similarity of the present results to prior research (Aguinis et al., 2011) suggests that it did not greatly bias our results.

It is also important to point out that the magnitude of a particular effect is not necessarily an indication of its importance. As Rosenthal (1994) stated, a small effect can have important real-world consequences. However, knowledge of the magnitude and variability of effects within a literature can be useful information for a variety of reasons; it provides a gauge against which researchers can compare effect sizes across literatures and with their own research. This information also aids in statistical power analysis which increases the likelihood of detecting statistically significant effects.

This is especially important because a notable finding of this report is that more than half of all meta-analyses are based on studies that are, on average, statistically underpowered. On one hand, this finding is somewhat damning of the field as whole in that research study design issues still seem to be a major problem in management research and related fields despite some indications that researchers are paying more attention to critical issues like sample size (see Scandura & Williams, 2000). On the other hand, this finding also illustrates the importance of meta-analytic reviews (see Chan & Arvey, 2012, for an excellent summary of the merits of meta-analytic reviews). This review itself demonstrates that many, if not most, of the primary studies conducted in the management literature are not properly designed to accurately detect the effects of interest. Consequently, the chances of reporting inaccurate findings are unnecessarily high. However, because meta-analyses overcome the lack of power in individual studies, researchers are better able to demonstrate effects when they do exist. As Asendorpf et al. (2013) concluded, we need to consistently replicate primary studies to provide the fodder for meta-analysis.

Acknowledgments

We gratefully acknowledge the assistance of the editor, Fred Luthans, and three anonymous reviewers in preparing this research for publication. We also would like to thank Herman Aguinis for his insightful and helpful comments on an earlier version of this article.

Authors' Note

Some of the findings reported in this article were presented at the 27th Annual Conference of the Society for Industrial and Organizational Psychology in San Diego, California.

Declaration of Conflicting Interests

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

Funding

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

References

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*(1), 94-106.

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, *37*(1), 5-38.

Aguinis, H., Werner, S., Abbott, J. A. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, *13*, 515-539.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Argyris, C. (1957). *Personality and organization: The conflict between system and the individual*. New York, NY: Harper & Row.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., . . . Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108-119.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*, 496-515.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543-554.

Bass, B. M. (1954). The leaderless group discussion. *Psychological Bulletin*, *51*, 465-492.

Bligh, M. C., Kohles, J. C., Pearce, C. L., Justin, J. E., & Stovall, J. F. (2007). When the romance is over: Follower perspectives of aversive leadership. *Journal of Applied Psychology*, *56*, 528-557.

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, *100*, 431-449.

Brewer, J. K. (1972). On the power of statistical tests in the "American Educational Research Journal." *American Educational Research Journal*, *9*, 391-401.

Campion, M. (1993). Article review checklist: A criterion checklist for reviewing research articles in applied psychology. *Personnel Psychology*, *46*, 705-718.

Cashen, L. H., & Geiger, S. W. (2004). Statistical power and the testing of null hypotheses: A review of contemporary management research and recommendations for future studies. *Organizational Research Methods*, *7*, 151-167.

Chan, M. L. E., & Arvey, R. (2012). Meta-analysis and the development of knowledge. *Perspectives on Psychological Science*, *7*(1), 79-92.

Chase, L. J., & Chase, R. B. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology*, *61*, 234-237.

Chase, L. J., & Tucker, R. K. (1975). A power-analytic examination of contemporary communication research. *Communications Monographs*, *42*(1), 29-41.

Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the annual conference of the British Educational Research Association, University of Exeter, *England*.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*, 145-153.

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

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology*, *65*, 221-249.

Dalton, D. R., Daily, C. M., Johnson, J. L., & Ellstrand, A. E. (1999). Number of directors and financial performance: A meta-analysis. *Academy of Management Journal*, *19*, 674-686.

Edwards, E. R. (2008). To prosper, organizational psychology should . . . overcome methodological barriers to progress. *Journal of Organizational Behavior*, *29*, 469-491.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532-538.

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, *17*(1), 120-128.

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*, 555-561.

Ferguson, T. D., & Ketchen, D. J., Jr. (1999). Organizational configurations and performance: The role of statistical power in extant research. *Strategic Management Journal*, *20*, 385-395.

Fisher, R. A. (1925). *Statistical methods for research workers*. London, England: Oliver & Boyd.

Fritz, A., Scherndl, T., & Kühberger, A. (2012). A comprehensive review of reporting practices in psychological journals: Are effect sizes really enough? *Theory & Psychology*, *23*(1), 98-122.

Glass, G. V., & Hakstian, A. R. (1969). Measures of association in comparative experiments: Their development and interpretation. *American Educational Research Journal*, *6*, 403-414.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research* (*Vol. 56*). Beverly Hills, CA: Sage.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*(1), 1-20.

Greenwald, A. G. (1993). Consequences of prejudice against the null hypothesis. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences* (pp. 419-448). Hillsdale, NJ: Lawrence Erlbaum.

Hall, R., de Antueno, C., & Webber, A. (2007). Publication bias in the medical literature: A review by a Canadian Research Ethics Board. *Canadian Journal of Anesthesia*, *54*, 380-388.

Hall, S. M., & Brannick, M. T. (2002). Comparison of two random-effects methods of meta-analysis. *Journal of Applied Psychology*, *87*, 377-389.

Howard, G. S., Maxwell, S. E., & Fleming, K. J. (2000). The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods*, *5*, 315-332.

Hunter, J. E., & Hirsh, H. R. (1987). Applications of meta-analysis. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 321-357). New York, NY: Wiley.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Jenkins, W. O. (1947). A review of leadership studies with particular reference to military problems. *Psychological Bulletin*, *44*(1), 54-79.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, *127*, 376-407.

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, *15*, 624-662.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746-759.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254-284.

Kroll, R. M., & Chase, L. J. (1975). Communication disorders: A power analytic assessment of recent research. *Journal of Communication Disorders*, *8*, 237-247.

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against non-significant findings. *Communication Monographs*, *76*, 286-302.

Lipsey, M. W. (Ed.). (1998). *Design sensitivity: Statistical power for applied experimental research*. Thousand Oaks, CA: Sage.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181-1209.

Maxwell, S. E. (2003). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, *9*, 147-163.

Mazen, A. M., Graf, L. A., Kellogg, C. E., & Hemmasi, M. (1987). Statistical power in contemporary management research. *Academy of Management Journal*, *30*, 369-380.

Mazen, A. M., Hemmasi, M., & Lewis, M. F. (2006). Assessment of statistical power in contemporary strategy research. *Strategic Management Journal*, *8*, 403-410.

Meindl, J. R., Ehrlich, S. B., & Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly*, *30*(1), 78-102.

Mone, M. A., Mueller, G. C., & Mauland, W. (1996). The perceptions and usage of statistical power in applied psychology and management research. *Personnel Psychology*, *49*, 103-120.

Penick, J. E., & Brewer, J. K. (1972). The power of statistical tests in science teaching research. *Journal of Research in Science Teaching*, *9*, 377-381.

Peters, C. C., & Van Voorhis, W. R. (1940). *Statistical procedures and their mathematical bases*. New York, NY: McGraw-Hill.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, *112*, 160-164.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*, 331-363.

Rosenthal, R. (Ed.). (1994). *Parametric measures of effect size*. New York, NY: Russell Sage Foundation.

Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, *9*, 395-396.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*, 166-169.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical psychology*, *58*, 646-656.

Rothstein, H. R., & Bushman, B. J. (2012). Publication bias in psychological science: Comment on Ferguson and Brannick 2012. *Psychological Methods*, *17*, 129-136.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis*. West Sussex, England: John Wiley.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, *57*, 416-428.

Sawyer, A. G., & Ball, A. D. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Research*, *18*, 275-290.

Scandura, T., & Williams, E. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, *43*, 1248-1264.

Schmidt, F. L. (2008). Meta-Analysis: A constantly evolving research integration tool. *Organizational Research Methods*, *11*(1), 96-113.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hogrefe & Huber.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309-316.

Smith, H. L., & Krueger, L. M. (1933). A brief summary of literature on leadership. *Bulletin of Scholastic Education*, *9*(4).

Song, F., Parekh-Bhurke, S., Hooper, L., Loke, Y., Ryder, J., Sutton, A., . . . Harvey, I. (2009). Extent of publication bias in different categories of research cohorts: A meta-analysis of empirical studies. *BMC Medical Research Methodology*, *9*(1), 79. doi:10.1186/1471-2288-9-79

Steel, P. (2013). *Converting between meta-analytic variance estimates based on correlations and Fisher z*. Paper presented at the 8th Annual Meeting of the Society for Research Synthesis Methodology, *Providence, RI*.

Steel, P. D., & Kammeyer-Mueller, J. (2008). Bayesian variance estimation for meta-analysis: Quantifying our uncertainty. *Organizational Research Methods*, *11*(1), 54-78.

Steel, P. D., Kammeyer-Mueller, J., & Nguyen, B. (2012). Why meta-analysis exaggerates generalizability and underestimates situational specificity. In S. Morris (Chair), *Methodological advances in meta-analysis*. Symposium conducted at the annual meeting of the Society for Industrial Organizational Psychology, *San Diego, CA*.

Steiner, D. D., Lane, I. M., Dobbins, G. H., Schnur, A., & McConnell, S. (1991). A review of meta-analyses in organizational behavior and human resources management: An empirical assessment. *Educational and Psychological Measurement*, *51*, 609-626.

Stemig, M. S., & Sackett, P. R. (2013). *Another look into the file drawer problem in meta-analysis*. Paper presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, *Houston, TX*.

Stogdill, R. M. (1948). Personal factors associated with leadership: A survey of the literature. *Journal of Psychology*, *25*(1), 35-71.

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435-454). New York, NY: Russell Sage Foundation.

Tett, R. P., Meyer, J. P., & Roese, N. J. (1994). Applications of meta-analysis: 1987-1992. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (*Vol. 9*, pp. 71-112). London, England: John Wiley.

Weber, R., Camerer, C., Rottenstreich, Y., & Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization Science*, *12*, 582-598.

Yates, F. (1951). The influence of statistical methods for research workers on the development of the science of statistics. *Journal of the American Statistical Association*, *46*(253), 19-34.

Zedeck, S. (2003). Editorial. *Journal of Applied Psychology*, *88*, 3-5.

Author Biographies

Ted A. Paterson is an assistant professor of management at Oregon State University. He received his PhD from the University of Nebraska–Lincoln. His research interests include positive organizational behavior, leadership, identity, and ethics.

Peter D. Harms is an assistant professor of management at the University of Alabama. He received his PhD from the University of Illinois at Urbana-Champaign. His research interests include the assessment and development of personality, psychological well-being, and leadership.

Piers Steel is a professor in Human Resources and Organizational Behavior at the Haskayne School of Business and is the Distinguished Research Chair at the Canadian Centre for Advanced Leadership in Business. He is a recognized authority on the science of motivation and is known internationally for his procrastination research.

Marcus Credé is an assistant professor in psychology at Iowa State University, who obtained a PhD in industrial and organizational psychology from the University of Illinois at Urbana-Champaign in 2005. His research interests are focused on job attitudes and other noncognitive influences on performance as well as methodological threats to validity.