Journal of Management
Vol. XX No. X, Month XXXX 1–28
DOI: 10.1177/0149206317710723

Advancing Theory by Assessing Boundary Conditions With Metaregression: A Critical Review and Best-Practice Recommendations

Erik Gonzalez-Mulé
Indiana University
Herman Aguinis
George Washington University

Understanding boundary conditions, or situations when relations between variables change depending on values of other variables, is critical for theory advancement and for providing guidance for practice. Metaregression is ideally suited to investigate boundary conditions because it provides information on the presence and strength of such conditions. In spite of its potential, results of our review of 63 metaregression articles published in the Journal of Management, Journal of Applied Psychology, Personnel Psychology, Journal of Management, Academy of Management Journal, *and* Strategic Management Journal *uncovered a surprising lack of transparency, frequently implemented erroneous practices, and a lack of attention to important methodological choices. Results also suggest that many substantive conclusions are ambiguous at best and, unbeknownst to authors and readers, potentially misleading. Drawing from our review of the substantive literature as well as the latest statistical and methodological research, we offer evidence-based best-practice recommendations on how to conduct and report the results of a metaregression study. We offer recommendations on calculating statistical power and heterogeneity, choosing an appropriate model, testing boundary condition hypotheses, adjusting* R^2 *for known variance, explaining methodological choices, and reporting and interpreting model coefficients and other results. Also, we conducted two illustrative metaregression studies that incorporate all of our recommendations with accompanying syntax and data. Our recommendations can be used by authors, readers, journal editors, and reviewers wishing to*

Acknowledgments: We thank J. Craig Wallace and two Journal of Management *anonymous reviewers for their useful and constructive feedback, which allowed us to improve our manuscript substantially. Also, we thank Frank L. Schmidt for sharing his knowledge about metaregression with us.*

Supplemental material for this article is available with the manuscript on the JOM *website.*

Corresponding author: Erik Gonzalez-Mulé, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th St. HH3100, Bloomington, IN 47405, USA.

E-mail: erikgonz@indiana.edu

conduct and evaluate metaregression studies, as well as practitioners interested in understanding conditions under which organizational practices are more or less likely to be effective.

Keywords: *boundary conditions; meta-analysis; metaregression; research methods*

Boundary conditions specify situations when relations between variables change depending on values of other variables (Busse, Kach, & Wagner, 2016). Knowledge of boundary conditions is critical for theory advancement (Aguinis, Gottfredson, & Culpepper, 2013; Hall & Rosenthal, 1991; Mathieu, Aguinis, Culpepper, & Chen, 2012) because, for a theory to be useful, "the boundaries of the theory must be understood" (Bacharach, 1989: 498). From a practice perspective, theories that take into account boundary conditions provide more precise and nuanced guidance to organizational decision makers regarding the effectiveness of selection instruments (Barrick, Mount, & Li, 2013), training (Blume, Ford, Baldwin, & Huang, 2010) and performance feedback (Kluger & DeNisi, 1995) interventions, the design of jobs (Hackman & Oldham, 1975), and goal-setting programs (R. E. Wood, Mento, & Locke, 1987), among many other organizational practices and interventions.

Given their centrality for theory advancement and their usefulness for guiding practice, it is not surprising that research examining boundary conditions has been described in the pages of *Journal of Management* regularly (e.g., Aguinis, Gottfredson, & Culpepper, 2013; Boyd, Haynes, Hitt, Bergh, & Ketchen, 2012). This vast body of research has referred to boundary conditions with many labels, including *moderator variable* or *moderating effect* (Aguinis, 1995) and *interaction effect* (Aguinis, Gottfredson, & Culpepper, 2013). We therefore use the terms *boundary conditions*, *moderators*, and *interaction effects* interchangeably throughout our article. Underlying all of these terms is the same notion that there are contingent factors that change the nature and/or direction of relations among variables (Boyd et al., 2012). As such, although knowledge of boundary conditions improves precision and is therefore important to advance all theories, they constitute the core foundation of contingent and interactionist models in domains such as leadership (e.g., Fiedler, 1967), goal setting (e.g., Locke & Latham, 1990), personality (e.g., Barrick et al., 2013), and the prediction of individual (Van Iddekinge, Aguinis, Mackey, & DeOrtentiis, 2017) and firm (e.g., Boyd et al., 2012) performance, among others. In addition, the topic has received considerable attention from a methodological perspective (e.g., Aguinis, 1995; Aguinis, Gottfredson, & Culpepper, 2013; Boyd et al., 2012). After all, the accuracy of the assessment of boundary conditions is only as good as the methodological tools that we have at our disposal.

Thousands of individual studies have been conducted to assess boundary conditions (e.g., Aguinis, Beaty, Boik, & Pierce, 2005; Boyd et al., 2012). But, as noted by Hunter, Schmidt, and Jackson (1982: 10),

> 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.

Accordingly, since the 1980s (Cortina, Aguinis, & DeShon, 2017), management researchers have used meta-analysis to aggregate results from many primary studies and generate more accurate population-level estimates of the presence of boundary conditions. In fact, this meta-analytic literature has covered virtually all management domains, ranging from the prediction of turnover (Griffeth, Hom, & Gaertner, 2000) to the effects of diversity on workgroups (Webber & Donahue, 2001) and CEO compensation (van Essen, Otten, & Carberry, 2015).

In spite of their important contributions, traditional meta-analytic methods to assess boundary conditions include two important limitations. First, meta-analysts have historically examined boundary conditions using a subgrouping method, whereby the studies included in the meta-analysis are separated according to the hypothesized boundary conditions and meta-analytic effect size estimates are computed separately on these subsamples. This subgrouping method forces researchers to artificially dichotomize boundary conditions that, in many cases, are continuous in nature (Aguinis, Gottfredson, & Wright, 2011). For example, consider the relation between cognitive ability and performance and the potential boundary condition motivation such that we predict that the ability-performance relation will become stronger as motivation increases (Van Iddekinge et al., 2017). The subgrouping strategy allows only for a test of the ability-performance relation for "low" versus "high" motivation conditions. Thus, it leads to loss of information and, in many cases, the incorrect conclusion that the boundary condition does not exist (Aguinis & Stone-Romero, 1997). A second limitation of traditional meta-analysis is that it allows only for an examination of each boundary condition in isolation. The underlying assumption, which is likely to be violated in virtually all cases, is that the presence of a particular boundary condition is unaffected by the presence of other boundary conditions (Steel & Kammeyer-Mueller, 2002).

In an effort to overcome the aforementioned limitations of traditional meta-analysis to assess boundary conditions, researchers have more recently turned to weighted least-squares metaregression. Metaregression is similar to but also different from traditional regression models. Specifically, it involves regressing meta-analytically derived effect sizes on several hypothesized predictors (i.e., boundary conditions) while weighting each effect size estimate by an indicator of its precision, such as the inverse of the within-study variance (i.e., sampling variance) of the study. The first important benefit of metaregression is that it capitalizes on a major advantage of meta-analysis, which involves relying on all of the available data to examine a particular phenomenon. Second, metaregression avoids (a) artificial dichotomization of boundary conditions and (b) the untenable assumption that reality includes only one boundary condition at a time. Studies based on metaregression have already provided important theoretical contributions in various topical areas, such as resource dependence theory (Drees & Heugens, 2013), counterproductive behaviors (Gonzalez-Mulé, Mount, & Oh, 2014), transfer of training (Blume et al., 2010), and turnover (Griffeth et al., 2000).

As is usually the case with the adoption of novel methodological tools, initial usage involves some suboptimal practices, as has been documented regarding the adoption of meta-analytic methods (Schmidt & Hunter, 1999), longitudinal research (Ployhart & Vandenberg, 2010), and multilevel analysis (Aguinis, Gottfredson, & Culpepper, 2013). In fact, we reviewed 63 metaregression studies that were published in the *Journal of Management*, *Journal of Applied Psychology*, *Personnel Psychology*, *Academy of Management Journal*, and *Strategic Management Journal*, and we found a surprising lack of transparency, frequently implemented erroneous practices, and a lack of attention to important methodological choices. Most of these studies (n = 32, 51%) were published in 2010 or after, which indicates that metaregression is moving from a niche methodological tool to one that is an integral part of mainstream management research. The existing confusion regarding the execution and reporting of metaregression studies is not just a subtle issue of concern to "methodological experts" only. To the contrary, the lack of standards and inconsistent assessment of boundary conditions via metaregression form a timely topic in management research given the current debate about transparency and replicability of research results (Bergh, Sharp, Aguinis, & Li, 2017; Bosco, Aguinis, Field, Pierce, & Dalton, 2016; Kepes & McDaniel, 2013; O'Boyle, Banks, & Gonzalez-Mulé, 2017; Wright, 2016). As we illustrate later through articles based on our literature review, methodological choices in metaregression studies have significant consequences for substantive conclusions, making these choices of primary interest for substantive researchers whose goal is to advance theory.

Based on the aforementioned discussion, the goal of our article is to offer a critical review and evidence-based best-practice recommendations on how to assess boundary conditions with metaregression. We follow a structure similar to that of other methodological best-practice articles published in the *Journal of Management* (e.g., Aguinis, Gottfredson, & Culpepper, 2013; Aguinis, Forcum, & Joo, 2013; Boyd et al., 2012; Geyskens, Krishnan, Steenkamp, & Cunha, 2009; Richard, Devinney, Yip, & Johnson, 2009; Stromeyer, Miller, Sriramachandramurthy, & DeMartino, 2015; Zyphur & Oswald, 2015), as these recommendations can be used as a roadmap by authors, readers, and journal editors and reviewers wishing to conduct and evaluate metaregression studies. They can also be used by practitioners who plan on implementing interventions and practices based on the knowledge generated by metaregression research.

Our primary targeted audience includes substantive researchers with the usual doctoral-level methodological training in management and related fields, rather than expert methodologists. In an effort to reach this audience, we also conducted two illustrative metaregression studies that, together, offer specific and concrete guidance regarding the implementation of all of our best-practice recommendations. Furthermore, although many of the recommendations that we advance may be known by methodologists and substantive researchers in other fields, our review of the substantive literature suggests that our recommendations are not known in management and related fields. This likely stems from the absence of a single source that explains the various methodological issues and decision points that influence substantive conclusions in a relatively nontechnical format. Thus, the primary contribution of our article is to offer metaregression best practices to substantive researchers in management and related fields.

Next, we provide a brief overview of metaregression. Then, we offer best-practice recommendations based on a review of methodological articles in medicine (e.g., Higgins & Thompson, 2002), psychology (e.g., Overton, 1998), and ecology (e.g., Gurevitch & Hedges, 1999), as well as substantive journals in management and related fields. In many instances, the appropriate methodological choice may carry with it certain tradeoffs. In these cases, our goal is to draw attention to the different choices available and the situations when one choice might be more appropriate than another, while promoting greater transparency in reporting. Moreover, we highlight situations when use of metaregression may not be particularly useful or informative. Following our best-practice recommendations, we present two illustrative metaregression studies with a detailed, step-by-step explanation of the implementation of all of our recommendations as well as syntax (and the raw data) to reproduce our analyses. In short, our article offers a single source of evidence-based recommendations and a how-to guide regarding the conduct and reporting of a metaregression study.

Metaregression: A Brief Overview

The following equation (e.g., Raudenbush, 2009; Viechtbauer, López-López, Sánchez-Meca, & Marín-Martínez, 2015) describes the extent to which boundary conditions x predict the relation between two variables, as indexed by the effect size y, in a sample of k studies:

$$y_i = B_0 + B_1 x_{i1} + \cdots B_j x_{ij} + u_i + e_i, \tag{1}$$

where y_i denotes the effect size estimate (e.g., r's or d's) in the ith study; x_{ij} denotes the value of the jth boundary condition in the ith study; B_j denotes the unstandardized regression coefficient associated with boundary condition j that indicates the extent to which the effect size y changes as a result of a one-unit change in x_{ij}; and B_0 is the model intercept. In addition, u_i denotes the random-effects variance components with distribution $N(0, \tau^2_{\text{res}})$, and e_i denotes the within-study variances with distribution $N(0, v_i)$. Furthermore, τ^2_{res} denotes the residual heterogeneity, or the variability in the population of effect sizes not accounted for by the predictors (i.e., boundary conditions) in the model, and v_i represents the within-study variance of the studies.

Boundary conditions (i.e., x) are typically coded from information available in or extracted from the primary-level studies. For example, we may be interested in testing the hypothesis that the relation between cognitive ability and job performance is stronger in more complex jobs (Schmidt & Hunter, 2004). To do so, we would code the level of job complexity in each study, which would be the operationalization of x by, for example, assigning a score from 1 to 5 to each sample based on the jobs that the people in the sample occupy, with 1 indicating a lower level of complexity and 5, a high level of complexity.[1]

Ordinary least squares regression is one of the most frequently used data-analytic techniques in management and related fields (Aguinis, Pierce, Bosco, & Muslin, 2009). Thus, because ordinary least squares regression is a known point of reference, it is useful to compare and contrast it with metaregression. First, metaregression uses study-level, rather than primary-level, predictors. So, the predictors in Equation 1 are based on characteristics of the studies included in the meta-analysis. Note that, in contrast to the typical primary study assessing boundary conditions, the metaregression coefficients associated with the x variables provide information on the boundary conditions, and there is no need to include product terms in the equation. The reason is that the coefficients assess how much the relation between two variables, denoted by y in our model, changes as a function of x. In other words, positive (negative) coefficients indicate an increase (decrease) in effect size units with increasingly large values of x. Second, the criterion in metaregression is the effect size of the relation between two variables from primary studies rather than scores on the criterion. Thus, whereas in ordinary least squares regression, the criterion consists of a column of scores for a particular variable (e.g., job performance), in metaregression it is a column of effect size estimates, such as r's or d's (i.e., the relation between cognitive ability and job performance). Third, metaregression assigns a weight to each study to account for its precision, with larger studies given larger weight. This is an important feature of metaregression because it allows researchers to give more influence to studies providing more stable population estimates (i.e., those based on larger sample sizes). Fourth, the researcher must make several methodological choices and statistical adjustments that are nonexistent in more traditional regression models (Borenstein, Hedges, Higgins, & Rothstein, 2009), such as deciding whether to pursue metaregression analyses given the heterogeneity present in the studies as well as whether to adopt a fixed-effects (FE) or mixed-effects (ME) model,[2] and these choices have important implications for substantive conclusions.

Evidence-Based Best-Practice Recommendations for Using Metaregression

We present recommendations based on a review of the methodological and substantive literatures with the goal of highlighting evidence-based best practices. Specifically, we reviewed the methodological literature with the goal of identifying key decision points, and we coded the frequency with which authors made these decisions in a review of the substantive literature. The key decision points that we coded were as follows: heterogeneity estimate, type of model, residual heterogeneity estimate, hypothesis testing method, and whether authors presented a correlation matrix and descriptive statistics of the boundary conditions included in their study. We limited our coding to these issues because there was no information available in the articles with respect to other methodological choices that we later describe (e.g., conducting a power analysis). With respect to the substantive literature, we used results of a review and analysis of 63 articles that used metaregression and were published between 1982 and 2016 in the *Journal of Management* (n = 14), *Journal of Applied Psychology* (n = 31), *Personnel Psychology* (n = 10), *Academy of Management Journal* (n = 7), and *Strategic Management Journal* (n = 1). We include the list of 63 articles in the review, as well as information on article search procedures and inclusion criteria, in the supplemental online materials (see Appendix A).

As shown in Table 1, 43 studies (67%) did not report whether they used an FE or ME model, and 31 studies (49%) reported only weighting studies by an index of within-study variance. This implies an FE model, given that these studies did not state that they also weighted studies by an estimate of residual heterogeneity. Nine studies (17%) reported using an iterative index of residual heterogeneity, and 20 (32%) did not report the within-study variance or residual heterogeneity estimator used. Of the 12 (18%) studies that reported weighing by an index of residual heterogeneity, nine used the maximum likelihood or restricted maximum likelihood method. Furthermore, 10 studies (16%) did not include any information on the degree of heterogeneity present in their sample of studies; 29 (46%) reported only the statistical significance of heterogeneity; 59 (93%) did not describe procedures used for testing boundary condition hypotheses (i.e., hypothesis-testing method for regression coefficients); and 20 (32%) did not provide any information about the method used to estimate the residual heterogeneity or the within-study variance index used to weigh the studies. Furthermore, 49 studies (78%) did not report the intercorrelations among the boundary conditions. Note that power analysis is not included in Table 1, because not a single study among the 63 articles reported conducting one.[3]

In short, results of our review indicate that the substantive literature is remarkably opaque, with few studies reporting the necessary information for readers and reviewers to be able to gauge the choices that were made, their justification, and their suitability for each research context. Next, we describe our best-practice recommendations, which are organized around the conduct and reporting of results of a study assessing boundary conditions with metaregression. Our recommendations focus on issues specific to metaregression and not on additional decision points that relate to all types of meta-analysis more generally.[4] We base our recommendations

Table 1
Methodological Choices Regarding the Assessment of Boundary Conditions Based on Metaregression From Select Articles

Methodological Choices	Articles, *n* (%)
Heterogeneity estimate[a]	
Q statistic	29 (46)
Credibility interval	26 (41)
T^2	9 (14)
I^2	2 (3)
Not reported	10 (16)
Type of metaregression model[b]	
Fixed effects	2 (3)
Mixed effects	19 (30)
Not reported	43 (67)
Residual heterogeneity estimate	
Only within-study variance weight reported	31 (49)
Restricted maximum likelihood	9 (14)
Noniterative estimator of residual variance	3 (5)
Not reported	20 (32)
Boundary condition hypothesis-testing method	
Huber-White	1 (2)
Other	3 (5)
Not reported	59 (93)
Variable intercorrelations and descriptive statistics	
Reported	14 (22)
Not reported	49 (78)

Note: From articles in the *Journal of Management, Journal of Applied Psychology, Personnel Psychology, Academy of Management Journal*, and *Strategic Management Journal* (1982-2016). Total number of articles, $N = 63$ (the list is included in the online supplemental material; see Appendix A). The Q statistic refers to the chi-square significance test of heterogeneity; T^2 is an estimate of the population parameter τ^2 and refers to the between-studies variance; and I^2 is the ratio of heterogeneity to total variability.
[a]Because several studies reported multiple indices of heterogeneity, the percentages do not add to 100.
[b]Because one study reported both fixed- and mixed-effects results, the percentages do not add to 100.

on the empirical and analytical evidence available to date, and they are supported by results of simulation studies, analytic work, or both. When the available evidence does not clearly specify one correct way to proceed, we describe possible courses of action and the advantages and limitations associated with each alternative. Our best-practice recommendations, which we discuss in detail next, are summarized in Table 2, with exemplary illustrations from the substantive literature. Note that none of the articles that we reviewed adopted all of our recommendations. Therefore, the articles included in Table 2 are exemplars only with respect to the specific recommendation with which they are associated.

Conducting a Metaregression Study

Calculation of statistical power. Statistical power is the probability that existing population boundary conditions will be detected. The power of the overall model is the probability that at least one boundary condition will be detected, while the power of the individual coef-

Table 2

Best-Practice Recommendations for Conducting and Reporting Results of a Study Assessing Boundary Conditions Based on Metaregression

Issue	Recommendation	Examples of Implementation of Recommended Practice
	Conducting a metaregression study	
1. Calculation of statistical power		
• Power is the probability that existing boundary conditions will be detected. In cases of insufficient statistical power, substantive conclusions about the presence of boundary conditions may be incorrect because existing boundary conditions are unlikely to be detected.	• Calculate the statistical power for the overall model (i.e., probability that at least one existing boundary condition will be detected) and for each hypothesized boundary condition before using metaregression. • If power is found to be low, do not use metaregression or interpret results in the context of low power, as nonsignificant results could be Type 2 errors (i.e., false negatives).	None
2. Calculation of heterogeneity		
• Heterogeneity is the systemic variability in the effect sizes across primary-level studies. In cases of low heterogeneity, boundary conditions are unlikely to be present. • The presence of heterogeneity provides prima facie evidence to proceed with tests of boundary conditions.	• Calculate and report multiple indices of heterogeneity, and use several indices in deciding whether to proceed with metaregression analyses (see Table 3). • Compare typical estimates of heterogeneity (e.g., T^2) with estimates derived from Bayesian methods when informed priors are available. • If the evidence suggests little heterogeneity, metaregression is unlikely to lead to meaningful theory advancements.	Carney, Gedajlovic, Heugens, van Essen, and van Oosterhout (2011); Heavey, Holwerda, and Hausknecht (2013)
3. Choosing type of metaregression model		
• The two types of metaregression models (ME and FE) have a specific set of assumptions and computational considerations. This results in unique situations when their use is appropriate (see Table 4). • The incorrect choice of model can result in an increase in Type 1 error rates, a decrease in power, and incorrect effect size estimates.	• Use the ME model except for the specific instances where an FE model may be appropriate (see Table 4). • With either model, provide a justification for the choice of model and, if using an FE model, show ME results and provide evidence that the assumptions of the FE model have been met (see Table 4).	Drees and Heugens (2013); Heugens and Lander (2009); Rabl, Jayasinghe, Gerhart, and Kühlmann (2014)
4. Calculation of residual heterogeneity		
• Residual heterogeneity (i.e., τ^2_{res}) is the variability in effects not accounted for by the boundary conditions included in the model and can be estimated with either iterative or noniterative estimators. • A noniterative estimator increases Type 1 error rates, decreases power, and reduces the predictive power of the model.	• Use an iterative estimator of residual heterogeneity, such as the REML estimator. These estimators are more accurate than noniterative estimators, especially when the number of studies is small (i.e., $k < 40$). • The EB estimator can be used when there are sufficient data available to generate an informed prior distribution.	Park and Shaw (2013); Rabl, Jayasinghe, Gerhart, and Kühlmann (2014); Sturman, Cheramie, and Cashen (2005)
5. Testing boundary condition hypotheses		
• The popular Wald-type method yields incorrect standard errors and false conclusions regarding boundary condition hypotheses because it makes incorrect sample size adjustments and, in ME models, does not allow for uncertainty in the estimate of τ^2_{res}. • The Wald-type method results in Type 1 error rates up to four times the nominal level.	• Use the Hedges and Olkin (2014) method for FE models. • Use the Knapp and Hartung (2003) method for ME models. • Only use the robust standard errors produced by the HLM program (i.e., the Huber-White method) when $k > 80$.	Hüffmeier, Freund, Zerres, Backhaus, and Hertel (2014)

(continued)

Table 2 (continued)

Issue	Recommendation	Examples of Implementation of Recommended Practice
6. Adjusting R^2 for known variance		
• Estimates of variance explained (i.e., R^2) obtained from metaregression analyses are biased downwardly because R^2 treats within-study variance as unexplained variance, whereas in metaregression, within-study variance can be quantified.	• Calculate R^2_{Meta} to adjust R^2 such that it takes into account the known within-study variance in a metaregression context.	None
Reporting results of a metaregression study		
7. Explanation of methodological choices		
• It is important to be transparent with respect to all methodological choices. • Without transparency, there is a low likelihood that study findings can be reproduced and replicated, which contributes to a decrease in the credibility and legitimacy of study findings.	• Follow an overarching principle of transparency with respect to methodological reporting. • Explain and justify all methodological choices, including the type of model, the estimator of residual heterogeneity (for ME models), and the method used for statistical significance testing.	Drees and Heugens (2013); Heugens and Lander (2009); Park and Shaw (2013)
8. Reporting and interpreting meta-regression coefficients		
• Unstandardized regression coefficients can be used to communicate practical importance by calculating actual effect size values at different values of the boundary condition, while standardized regression coefficients provide a common metric of relative importance. • Reporting only one type of regression coefficient makes the practical significance of study results more difficult to interpret.	• Report unstandardized and standardized regression coefficients. • Use unstandardized regression coefficients from the metaregression analyses to generate values of the effect size at different theoretically or practically relevant values of the boundary condition under study.	Sturman (2003); Sturman, Cheramie, and Cashen (2005)
9. Reporting variable intercorrelations and descriptive statistics		
• Most journals require reporting descriptive statistics and variable intercorrelations for primary studies. • In the absence of this information, readers are unable to fully evaluate metaregression findings.	• Report the intercorrelations and descriptive statistics of the boundary conditions and effect sizes.	D'Innocenzo, Mathieu, and Kukenberger (2014); Hausknecht, Halpert, Di Paolo, and Moriarty Gerrard (2007)

Note: T^2, between-studies variance; ME, mixed-effects model; FE, fixed-effects model; τ^2_{res}, residual heterogeneity; REML, restricted maximum likelihood; EB, empirical Bayes; HLM, hierarchical linear modeling; R^2, percentage of variance explained by the regression equation.

ficients refers to the probability that each specific boundary condition will be detected. Insufficient power means that substantive conclusions about the presence of boundary conditions will be incorrect because existing boundary conditions are unlikely to be found. Hedges and Pigott (2004) highlighted power in metaregression as a function of several factors, including the estimation method used, the number of studies in the meta-analysis (i.e., the effective sample size for the metaregression equation), whether moderators are continuous or categorical, the degree of heterogeneity in the sample of studies, the a priori Type 1 error rate, the expected size of the boundary condition, and other factors. Hedges and Pigott also provided equations that can calculate power under different conditions.

We were unable to identify any articles that reported conducting a power analysis before performing metaregression analyses, although some studies reported using an ad hoc decision rule (e.g., minimum of 10 studies) to decide whether to use metaregression (e.g., Zhong, Su, Peng, & Yang, 2014), possibly because the formulae presented in Hedges and Pigott (2004) are complex and require an understanding of matrix algebra to calculate power. Fortunately, Cafri, Kromrey, and Brannick (2009) developed an SAS macro that calculates power of the overall metaregression equation and individual boundary conditions with the formulae developed by Hedges and Pigott (2004).[5]

Using the Cafri et al. (2009) macro, we conducted a simulation using 1,000 samples with input values reflecting the median values that we gathered from the studies in our review in terms of number of studies ($k = 37$), sample size per study ($N = 150$), and reasonable values for other parameters, such as a medium expected moderator effect size of .16 (Bosco, Aguinis, Singh, Field, & Pierce, 2015) and a medium heterogeneity estimate of .02 (Hedges & Pigott, 2004) required for power calculation. We used the power calculator on these simulated data sets and found that the average power across 1,000 simulated samples was .66 for FE models and .40 for ME models with one moderator so that the power of the overall model and individual coefficients is equivalent. Both values are closer to the result of a coin flip than to Cohen's (1992) frequently cited recommended power level of at least .80. Given that many studies likely have insufficient statistical power, it is likely that researchers erroneously conclude that there are no statistically significant boundary conditions in their metaregression analyses.

In sum, we recommend that authors calculate the overall statistical power of their metaregression model and the power to detect each hypothesized boundary condition before using metaregression and that they report the results of this analysis in all metaregression studies. We also recommend that if power is found to be low, metaregression should not be conducted or, at the very least, the results and conclusions of the study should be interpreted in the context of low power because nonsignificant results could be Type 2 errors (i.e., false negatives).

Calculation of heterogeneity. Heterogeneity is the systemic variability (i.e., not attributable to sampling error) in the primary-level effect sizes. A large degree of heterogeneity indicates that boundary conditions are likely present, and little heterogeneity indicates that boundary conditions are unlikely to be present (Aguinis & Pierce, 1998; Higgins & Thompson, 2002). Thus, tests of heterogeneity provide prima facie evidence regarding the potential usefulness of investigating boundary conditions. In Table 3, we provide a summary of methods available to estimate heterogeneity, with a description of the advantages and disadvantages of each.

First, the Q statistic is the weighted residual sum of squares between individual study effect sizes and the mean effect size across studies. The Q statistic follows a chi-square distribution and can be used to test the null hypothesis that heterogeneity is zero in the population. Q's statistical significance provides an easy-to-use decision rule regarding whether to proceed with tests of boundary conditions (Hedges & Olkin, 2014). However, the Q statistic does not provide an estimate of the amount of heterogeneity, and its statistical significance is largely based on sample size. Furthermore, because of its underlying basis in statistical significance testing, the Q statistic is underpowered with too few studies and will also detect a

Table 3
Comparison of Methods Used to Calculate Heterogeneity in Meta-Analysis

Heterogeneity Method	Advantages	Disadvantages
Q statistic: weighted residual sum of squares between individual study effect sizes and the mean effect size across studies.	• Can be tested for statistical significance • Provides easy-to-use decision rule	• Does not provide estimate of the magnitude of heterogeneity • Underpowered with too few studies and detects a trivial amount of heterogeneity with many studies
Credibility interval: estimate of variability around the population effect size, after within-study variance and the effects of other methodological and statistical artifacts have been removed.	• Absolute metric for evaluating the presence of heterogeneity • Provides upper and lower bounds of population effect size	• Unclear guidelines on what constitutes a narrow or wide credibility interval • Estimate of variability used to calculate credibility interval criticized as being an underestimate
T^2: estimate of the parameter τ^2 and denotes the between-study variance, computed as the Q statistic minus the degrees of freedom (i.e., $k - 1$) divided by a scaling factor.	• Absolute metric for evaluating the presence of heterogeneity • Provides direct estimate of the magnitude of heterogeneity • Can be used to calculate credibility intervals around population estimate	• Unclear guidelines on what constitutes a small or large value of T^2 • Cannot be used to compare heterogeneity across meta-analyses
I^2: ratio of between-study variance to total variance.	• Relative metric allows for comparing the degree of heterogeneity across studies • Easily interpretable as the percentage of variance attributable to between-study effects	• Does not provide information regarding the absolute amount of heterogeneity
Bayesian method: uses informed prior based on extant meta-analytic evidence to generate probability distribution of heterogeneity values	• Provides more accurate estimate of heterogeneity than T^2 • More robust to small sample sizes	• Requires the use of informed prior, which may not be available

trivial amount of heterogeneity with many studies. This increases the likelihood that a researcher will identify a statistically significant boundary condition that may have little importance for theory or practice (Borenstein et al., 2009; Higgins & Thompson, 2002).

Second, the credibility interval (CrI) provides an estimate of the variability around the population effect size, after within-study variance and other study artifacts have been removed (Schmidt & Hunter, 2014). The CrI provides an absolute metric for evaluating the presence of heterogeneity, as it is on the same scale as the effect size (e.g., r or d). Furthermore, the CrI provides upper and lower bounds of the population effect size, with the typical decision rule being that if the CrI is sufficiently wide or includes zero, tests for boundary conditions are justified (Whitener, 1990). However, there is disagreement over what constitutes a narrow or wide CrI, and some have criticized the estimate of variability in the true effects used to calculate the CrI as underestimating heterogeneity (Erez, Bloom, & Wells, 1996).

Third, the T^2 statistic is an estimate of the parameter τ^2 and denotes the between-study variance, computed as the Q statistic minus the degrees of freedom (i.e., $k - 1$) divided by a scaling factor (Borenstein et al., 2009). Thus, T^2 provides a direct estimate of the magnitude of heterogeneity. Like CrI, T^2 provides an absolute metric of heterogeneity, as it is on the same metric as the effect size (e.g., *r* or *d*), and its square root (i.e., *T*) can be used to compute a credibility interval around the population effect size. If the credibility interval is sufficiently wide, researchers can proceed with their boundary condition tests. However, like the CrI, it is unclear what constitutes a small or large value of T^2, and it cannot be used to compare heterogeneity across meta-analyses, as it is a local estimate of heterogeneity.

Fourth, the I^2 statistic is a relative metric of heterogeneity (Higgins & Thompson, 2002). Specifically, I^2 is the ratio of between-study variance to total variance. While CrI and T^2 are on the same scale as the effect size being analyzed in a meta-analysis, I^2 is on a scale of 0 to 1. Thus, although it does not provide information regarding the absolute amount of heterogeneity, the I^2 statistic can be used to compare the proportion of heterogeneity across different meta-analyses by interpreting it as the percentage of variance attributable to between-study effects. Higgins and Thompson (2002) concluded that I^2 values >.50 (i.e., >50% of the variance in effect sizes is attributable to between-study effects) indicate heterogeneity that should be explored further.

Finally, Steel, Kammeyer-Mueller, and Paterson (2015) recommended using a Bayesian method to estimate heterogeneity. Specifically, their method requires the specification of informed priors based on extant meta-analytic evidence and then the use of these data to build a probability density distribution of heterogeneity values. In this distribution, more likely values of heterogeneity (as determined by the informed priors) are given greater weight. On the basis of a Monte Carlo simulation and a reanalysis of several published studies, they concluded that frequentist methods (e.g., T^2) underestimate heterogeneity and that the Bayesian method allows for more accurate estimation, particularly when the number of studies is small. A challenge in all Bayesian analysis is the computation of a prior distribution (Zyphur & Oswald, 2015). Fortunately, priors in the context of metaregression can be created with the database of about 150,000 effect sizes made available by Bosco et al. (2015).

Borenstein et al. (2009) and others (e.g., Geyskens et al., 2009; Higgins & Thompson, 2002) suggested reporting several estimates of heterogeneity that provide evidence regarding the statistical significance (e.g., the *Q* statistic) as well as the absolute (e.g., CrI, T^2) and relative (e.g., I^2) degree of heterogeneity because presenting only one estimate of heterogeneity paints an incomplete picture. For example, Heavey, Holwerda, and Hausknecht (2013) conducted a meta-analysis on the causes and consequences of collective turnover; reported several indices of heterogeneity, including the *Q* statistic, the T^2 statistic, and the I^2 statistic; and used them in concert to justify their decision to examine boundary conditions. This study and a study by Carney, Gedajlovic, Heugens, van Essen, and van Oosterhout (2011) were the only two studies out of the 63 in our review that reported I^2. Furthermore, no studies used a Bayesian method to calculate heterogeneity as part of metaregression analyses, which is likely attributable to the infancy of Bayesian methods in management (Zyphur & Oswald, 2015).

In sum, in terms of the calculation of heterogeneity, we recommend that authors calculate, use, and report multiple indices of heterogeneity (i.e., the statistical significance of the heterogeneity, the absolute amount of heterogeneity, and the relative amount of heterogeneity) in deciding whether to proceed with metaregression analyses. We also suggest comparing typical heterogeneity estimates (e.g., T^2) to estimates derived from Bayesian methods (Steel et al., 2015) when the literature examined is sufficiently mature that informed priors are available. In the event that there is sufficient heterogeneity to justify the use of metaregression based on

Table 4
Comparison of Mixed- and Fixed-Effects Metaregression Models

Model	Key Assumptions	Computational Details	When Most Appropriate to Use
Mixed effects	• Sample of studies is a random draw from the population of studies. • Not every possible boundary condition is represented in the model.	• Studies weighted by residual heterogeneity and within-study variance • Standard errors are larger, thus allowing for greater uncertainty in the estimates. • Less weight is given to large sample studies as compared with the fixed-effects model.	• Additional boundary conditions not included in the model may exist. • The sample domain does not match the population domain. • Nascent field of study
Fixed effects	• Sample of studies is the population. • All boundary conditions are included in the model.	• Studies weighted by within-study variance only • More weight is given to large sample studies as compared with the mixed-effects model.	• All boundary conditions included in the model • The sample domain closely matches the population domain. • Mature domain of study

an examination of multiple indices, there should be an explanation interpreting the indices of heterogeneity and providing a rationale for proceeding with the analyses. If the body of evidence suggests little heterogeneity in the effect sizes, metaregression is unlikely to lead to meaningful theory advancements.

Choosing type of metaregression model. As is the case in meta-analysis more generally, there are two types of metaregression models: ME and FE. Most available meta-analysis programs with metaregression capabilities, such as Comprehensive Meta-Analysis (Borenstein et al., 2009), the metafor package in R (Viechtbauer, 2010), and Wilson's (2005) SPSS macros, can accommodate both types of models. As summarized in Table 4, each model has a specific set of assumptions, computational considerations, and situations when they should be used. Specifically, ME models assume that the sample of studies is a random draw from the population and that not every possible boundary condition is included in the model. However, FE models assume that the sample of studies *is* the population and that all boundary conditions are included in the model. The ME model is a combination of fixed effects (i.e., the boundary conditions hypothesized by the researcher) and random effects (i.e., the residual heterogeneity after accounting for moderators), while the FE model does not allow for residual heterogeneity. Thus, ME models have been conceptualized as a special case of multilevel models (e.g., Erez et al., 1996; Raudenbush, 2009), where individual study effect size estimates and associated within-study variance are treated as Level 1 variables and moderators are treated as Level 2 variables.

To compare the computational characteristics of the two models, the general weighting scheme for the studies analyzed in metaregression is described by the following equation:

$$w_i = 1/\left[v_i + \tau^2_{res}\right], \tag{2}$$

where w_i is the weight assigned to the *i*th study; v_i is the within-study variance of the *i*th study; and τ^2_{res} is an estimate of the residual heterogeneity after accounting for the moderators included in the model. In the FE model, τ^2_{res} is assumed to be zero, while in the ME model, τ^2_{res} is assumed to be an unknown parameter (we later discuss the methods available to estimate τ^2_{res} and the standard error of the regression coefficients). As Equation 2 shows, ME models allow for the residual heterogeneity and within-study variance to affect the weights of the effect sizes. This also affects the regression coefficients and results in larger standard errors, thus allowing for greater uncertainty in the estimates. This computational difference results in ME models assigning less weight to larger sample studies and more weight to smaller sample studies than FE models, which may result in effect size differences between the two (Borenstein et al., 2009).

The choice of whether to use the ME or FE model is a particularly important decision point because it affects Type 1 error rates, the power of statistical tests, and the effect size estimate (Erez et al., 1996; Overton, 1998; Viechtbauer et al., 2015). For example, a simulation study by Overton (1998) found that although FE models are more powerful than ME models to detect a true moderating effect, Type 1 error rates (i.e., false positives regarding the existence of boundary conditions) for FE models can exceed .50, or 10 times the usual .05 nominal level, when there is moderate heterogeneity. The high likelihood of false-positive results may lead to the incorrect conclusion that there is a boundary condition and therefore may provide inaccurate information to organizational decision makers. As Overton (1998: 376) concluded, FE models are only appropriate "when the contextual conditions are sufficiently defined *and* the sample domain closely matches the population domain." In other words, FE models are appropriate only in mature fields of study, when one is sufficiently confident that all possible moderators have been identified and included in the model and when the meta-analytic sample is sufficiently representative of the population; in contrast, ME models are appropriate in more nascent fields of study, when the assumptions of the FE model are not satisfied. The shortcomings associated with FE models have been documented in management and other fields (e.g., Erez et al., 1996; Gurevitch & Hedges, 1999; Thompson & Higgins, 2002).

Some studies in our review were transparent about the model they used, and used an ME model. For example, Drees and Heugens (2013) investigated the extent to which various institutional, field-level, and study characteristics moderated the relation between resource dependence and interorganizational arrangement formation and between interorganizational arrangement formation and firm performance. In their study, they outlined the relative benefits of using ME and FE models, and they presented their results with the ME model (see also Heugens & Lander, 2009).

In sum, in terms of choosing the type of metaregression model, our recommendation is to use the ME model except for the specific instances where an FE model may be appropriate. Those instances include a mature field of study where all boundary conditions are included in the metaregression model and when the sample of studies in the meta-analytic database is representative of the entire population of studies. When using either model, researchers should provide a justification for it and, if using an FE model, also show ME results and provide evidence that the assumptions of the FE model have been met (see Table 4).

Calculation of residual heterogeneity. When researchers choose an ME model, they must also decide how to estimate residual heterogeneity, τ^2_{res}. As we noted, τ^2_{res} is assumed to be zero in the FE model, and studies are weighted only by the inverse of the within-study variance (Borenstein et al., 2009). Residual heterogeneity is the variability in effects not accounted for by the boundary conditions included in the model (i.e., the variance that is "left over") and, in conjunction with the within-study variance (v_i in Equation 2), is used to weigh the studies in the metaregression analysis. Note that τ^2_{res} is a population parameter that cannot be calculated exactly, and any estimate of τ^2_{res} includes some degree of error. The choice of residual heterogeneity estimator has substantive consequences with respect to Type 1 error rates, statistical power, and the predictive validity of the model because different estimators have different distributional assumptions, weigh studies differently, and vary in terms of their computational demands (López-López, Marín-Martínez, Sánchez-Meca, Van den Noortgate, & Viechtbauer, 2014; Thompson & Sharp, 1999).

There are seven often-used residual heterogeneity estimators that can be classified into two groups: iterative and noniterative estimators. Both types generate an estimate of τ^2_{res} based on the heterogeneity in the studies and the covariance matrix of the proposed boundary conditions. Iterative estimators can be calculated with computer programs such as hierarchical linear modeling (Raudenbush, 2009), the metafor package in R (Viechtbauer, 2010), or macros designed for SPSS and other statistics software (Wilson, 2005). Simulation studies showed that iterative estimators, such as the restricted maximum likelihood and empirical Bayes estimators, perform best with respect to accuracy and bias, which indicate the extent to which the estimator generates estimates that deviate from the population parameter and the extent to which the estimator systematically over- or underestimates τ^2_{res}, respectively (e.g., López-López et al., 2014; Pocock, Cook, & Beresford, 1981; Sidik & Jonkman, 2007; Steel et al., 2015; Thompson & Sharp, 1999; Van den Noortgate & Onghena, 2003). This is especially the case when the number of studies is small (i.e., $k < 40$), with iterative estimators being significantly more accurate than noniterative estimates under these conditions. Several studies that we identified in our review used an iterative estimator. For example, Park and Shaw (2013) examined the extent to which entity size and average turnover rates moderated the relation between turnover and firm performance. They reported using a maximum likelihood estimator to estimate the residual heterogeneity in their regression models (see also Rabl, Jayasinghe, Gerhart, & Kühlmann, 2014; Sturman, Cheramie, & Cashen, 2005).

In sum, we recommend using an iterative estimator of residual heterogeneity, such as the restricted maximum likelihood estimator. This is especially important if the number of studies is small ($k < 40$). The empirical Bayes estimator can also be used when there is sufficient data available to generate an informed prior distribution of residual heterogeneity, for example, when conducting an updated meta-analysis to one that was published in the past.

Testing boundary condition hypotheses. When used to test boundary condition hypotheses with metaregression, the popular "Wald-type" method to estimate standard errors (i.e., the standard method used in weighted least-squares regression) yields incorrect estimates and false conclusions regarding the statistical significance of the regression coefficients. The reason is that the Wald-type method makes incorrect sample size adjustments to produce the standard errors of metaregression coefficients (Borenstein et al., 2009; Hedges & Olkin, 1985, 2014) and, in the context of ME models, the Wald-type method does not take into account uncertainty in the estimate of τ^2_{res} (Viechtbauer et al., 2015). Addressing this issue when using FE models is straightforward, as Hedges and Olkin (1985, 2014) recommended correcting the standard error by dividing it by the square root of the mean square residual and proceeding with hypothesis testing as normal with a z distribution. In ME models, the solution becomes more complex, as the method used to calculate the standard error of the regression coefficients has to take uncertainty in the estimate of τ^2_{res} into account. In fact, using the Wald-type method in an ME model results in Type 1 error rates up to four times the nominal level, depending on sample size, when the heterogeneity in the studies is greater than zero (Knapp & Hartung, 2003; Viechtbauer et al., 2015).

There are several alternative methods available to calculate the standard error of the regression coefficient to address this problem (e.g., Higgins & Thompson, 2004; Knapp & Hartung, 2003; Sidik & Jonkman, 2007). In a comprehensive simulation study, Viechtbauer et al. (2015) concluded that the Knapp and Hartung (2003) method provides the best statistical power and control of Type 1 error with an accompanying ease of computation over other similarly effective methods. Furthermore, the Wald-type method had the most inflated Type 1 error rates of the six examined by Viechtbauer et al. (2015). Several software programs, including the metafor package in R (Viechtbauer, 2010) and the Comprehensive Meta-Analysis computer program (Borenstein et al., 2009), have the capability to implement the Knapp and Hartung (2003) method.

In sum, in terms of testing boundary condition hypotheses, we recommend using the Hedges and Olkin (2014) method for FE models and the Knapp and Hartung (2003) method for ME models. Given that eight studies in our review used the hierarchical linear modeling program for their metaregression analyses, we also offer caution to researchers using this program and wishing to use the robust standard errors (i.e., the Huber-White method) to test the statistical significance of the boundary conditions in their study. Viechtbauer et al. (2015) found that the Huber-White method requires a large sample size ($k > 80$) to maintain a .05 Type 1 error rate, so we recommend using this method only when k is at least 80.

Adjusting R^2 for known variance. Estimates of variance explained (i.e., R^2) obtained from metaregression analyses are downwardly biased because R^2 in a typical regression context treats within-study variance as unexplained variance (Aloe, Becker, & Pigott, 2010). In metaregression, within-study variance can be quantified and used to calculate a more accurate estimate of R^2 as the percentage of between-studies variance explained by the model (i.e., R^2_{Meta}) in contrast to the typical method that calculates R^2 as the percentage of total variance accounted for by the model. The formula for R^2_{Meta} is the following:

$$R^2_{Meta} = 1 - \left[\tau^2_{res} / \tau^2\right], \tag{3}$$

where τ^2_{res} is the residual heterogeneity after accounting for the boundary conditions and τ^2 is the total heterogeneity. To calculate R^2_{Meta}, we use estimates of τ^2_{res} and τ^2 that we discussed previously. Aloe et al. (2010) concluded that typical regression models underestimate the amount of variance explained in the effect sizes by the boundary conditions included in the model. This is important because it suggests that metaregression models have more explanatory power than typically reported and that published metaregression studies actually underreported the explanatory power of theories. In our review of the substantive literature, we did not identify any studies that used Aloe and colleagues' method to calculate R^2_{Meta}, which leads us to believe that most studies underestimated the variance explained by boundary conditions.

In sum, we recommend that researchers calculate R^2_{Meta}. This adjustment takes into account the known within-study variance in a metaregression context.

Reporting Results of a Metaregression Study

In what follows, we describe several recommendations for reporting results of metaregression studies. The recommendations that we provide are made with the same overarching principle as that of Bernerth and Aguinis's (2016) recommendations for the use of statistical control variables: Researchers should report their choices in detail to ensure transparency and to maximize the likelihood of reproducibility of results in the future. This is based on the same rationale for the development of the Meta-Analytic Reporting Standards, as the American Psychological Association Publications and Communications Board stated that "without complete reporting of methods and results, the utility of studies for purposes of research synthesis and meta-analysis is diminished" (2008: 840).

Explanation of methodological choices. As noted earlier, a key issue that we identified in our review is authors' underreporting of key information regarding methodological choices. Thus, we recommend that researchers clearly explain and justify their methodological choices. For example, if a researcher wishes to use an FE model, there should be a clear explanation of why the FE model was used, including information about the heterogeneity of the studies, the maturity of the theory being examined, and the sampling population of studies (Overton, 1998). Similarly, if a researcher uses an ME model, there should be a clear rationale explaining why and an accompanying rationale to explain which estimator of residual heterogeneity was used (for an example, see Park & Shaw, 2013). In either case, researchers should also report the specific method used to test the statistical significance of the regression coefficients.

Reporting and interpreting regression coefficients. An important issue that we identified in our review is the lack of consistency regarding the reporting and interpretation of metaregression results. In fact, regression coefficients associated with a boundary condition are usually exclusively interpreted in terms of their statistical significance, as opposed to providing an interpretation that allows for an assessment of the magnitude as well as importance of the effect size for theory and practice. Unstandardized and standardized regression coefficients can both be used to convey this information.

Unstandardized coefficients are useful in terms of understanding practical importance because researchers can use unstandardized coefficients to generate meta-analytic effect size estimates under varying levels of the boundary condition. For example, Sturman et al. (2005) examined time as a boundary condition, with the expectation that the temporal consistency and stability of job performance ratings would decline as time between measurement episodes decreased. Therefore, they chose practically relevant time gaps and entered these values into their metaregression equation to provide actual estimates and accompanying confidence intervals of the temporal consistency and stability of job performance ratings at different time lags (see also Sturman, 2003). When there is no interest in particular values of boundary conditions, a usual choice is to select values that are one *SD* below and above the mean. However, because unstandardized coefficients are scale dependent, they cannot be used to compare the relative magnitude of effect sizes. Standardized coefficients allow for comparisons between boundary conditions within a metaregression model and with a separately conducted metaregression analysis. Thus, using standardized coefficients, researchers can compare the relative strength of one boundary condition over another. Reporting only one type of coefficient makes the practical significance of study results more difficult to interpret.

In sum, our recommendation is to report both unstandardized and standardized regression coefficients. This is consistent with the overarching theme of transparency that we discussed earlier, as both sets of coefficients provide important and complementary information. Furthermore, we recommend using the unstandardized regression coefficients from metaregression analyses to generate values of the effect size at different theoretically or practically relevant values of the boundary condition under study.

Reporting variable intercorrelations and descriptive statistics. Finally, most journals require reporting descriptive statistics (i.e., means, standard deviations) and variable intercorrelations in primary studies. In a metaregression context, reporting the intercorrelations between effect sizes and boundary conditions as well as descriptive statistics can help readers and reviewers better evaluate metaregression findings by identifying instances of high collinearity, redundancy between boundary conditions, or coding errors (Bedeian, 2013). For examples of the few studies that report this critical information, see D'Innocenzo, Mathieu, and Kukenberger (2014), and Hausknecht, Halpert, Di Paolo, and Moriarty Gerrard (2007). Thus, our final recommendation is to report the intercorrelations and descriptive statistics of the boundary conditions and effect sizes.

Implementing Best-Practice Recommendations: Two Illustrative Metaregression Studies

We now apply our recommendations to two illustrative metaregression studies. In our illustrations, we examine several moderators of the relations between (a) extraversion and job performance and (b) extraversion and job satisfaction using data from Gonzalez-Mulé (2015). In designing our study, we drew from the theory of purposeful work behavior (Barrick et al., 2013). In short, the theory argues that when the characteristics of people's jobs match their personalities, they will behave in more purposeful ways and enjoy enhanced perceptions of meaning at work (Aguinis & Glavas, 2017). The theory suggests that the relations of extraversion with job performance and job satisfaction will be stronger in jobs that involve leading others, jobs in which one's decisions have an impact on others, and jobs that involve contact with others. We coded boundary conditions using information available from the Occupational Information Network (O*NET) on a scale of 1 to 5, with higher numbers indicating that a higher level of the job characteristic is present in that job (as done by Judge & Zapata, 2015). Note that our focus is on the implementation of best-practice recommendations regarding metaregression, so we do not offer a description of theoretical background and hypotheses, nor do we discuss issues specific to the meta-analytic procedures. However, for details regarding data collection and other methodological procedures not specifically related to metaregression, see Appendix B in the online supplemental materials.

Table 5
Descriptive Statistics and Intercorrelations for Illustrative Metaregression Studies Examining Boundary Conditions of the Extraversion–Job Performance and Satisfaction Relations

	M	*SD*	1	2	3	4
1. Leading others	3.50 / 3.78	0.50 / 0.50		.63*	.31*	.15
2. Impact of decisions	3.95 / 3.91	0.48 / 0.61	.63*		.32*	.16
3. Contact with others	4.68 / 4.62	0.31 / 0.36	.31*	.32*		.24*
4. Effect size	.10 / .24	.17 / .22	.15	.16	.24*	

Note: Correlations for the extraversion–job performance relationship are below the diagonal, and the descriptive statistics are presented first; correlations for the extraversion–job satisfaction relationship are above the diagonal, and the descriptive statistics are presented second. Number of studies for the extraversion–job performance relationship, $k = 100$; for the extraversion–job satisfaction relationship, $k = 72$. Effect size refers to the individual correlations from each primary-level study.
*$p < .05$.

Table 5 shows the descriptive statistics and intercorrelations for our study variables. Our study characteristics are similar to those typically examined in the metaregression literature based on our review, with the exception that our sample sizes were somewhat larger ($k = 100$, $n = 15{,}667$ for job performance and $k = 72$, $n = 16{,}696$) than the median values from metaregression articles in the literature review that we described earlier in our discussion of power analyses. Nonetheless, our example is an illustration based on real data, even though we expect that the power to detect boundary conditions will be higher in our example than what is typical in the literature. In the following analyses, positive (negative) effect sizes associated with the moderators indicate that the relationships of extraversion with job performance and job satisfaction become more (less) strongly positive when the levels of these job characteristics are higher. As an additional value-added contribution of our article, all of the syntax that we used is available in the supplemental online materials (see Appendixes C and D), and the data are available at http://www.hermanaguinis.com. The full results for our metaregression analyses are reported in Table 6, which we refer to throughout our explanation.

Implementation of Recommendations

Our first recommendation regarding how to conduct a metaregression study refers to the calculation of statistical power. Thus, before conducting our analyses, we first calculated the statistical power of our metaregression equations. Using Hedges and Pigott's (2004) formulae in Cafri and colleagues' (2009) macro (for syntax, see Appendix C), we assumed a moderate sized effect of .16 and an alpha level of .05 (two-tailed). As shown in Table 6, for boundary conditions of the extraversion–job performance relation, we found that the power for the full model was 1.00 using both the FE and ME models. Tests to detect all of the individual coefficients had a power of 1.00, except for contact with others, which had a power of .91 in the ME model. As shown in Table 6, for boundary conditions of the extraversion–job satisfaction relation, power for the entire model was 1.00 based on the FE

Table 6
Metaregression Analyses of Boundary Conditions of the Extraversion-Job Performance and Satisfaction Relations

	Extraversion and Job Performance				Extraversion and Job Satisfaction			
	Fixed-Effects Model		Mixed-Effects Model		Fixed Effects-Model		Mixed-Effects Model	
Job Characteristic	*B*	β	*B*	β	*B*	β	*B*	β
Leading others (x_1)	.03† (.02, .07)	.11	.03 (.03, .41)	.09	.09* (.02, .00)	.22	.03 (.07, .71)	.06
Impact of decisions (x_2)	.05* (.02, .01)	.15	.04 (.04, .28)	.11	−.04* (.01, .01)	−.12	.02 (.06, .73)	.05
Contact with others (x_3)	.13* (.02, .00)	.32	.12* (.05, .01)	.24	.18* (.02, .00)	.31	.13† (.08, .09)	.21
R^2	.15		.10		.15		.06	
R^2_{Meta}	—		.12		—		.04	
Overall statistical power	1.00		1.00		1.00		1.00	
Statistical power of x_1	1.00		1.00		1.00		.67	
Statistical power of x_2	1.00		1.00		1.00		.84	
Statistical power of x_3	1.00		.91		1.00		.57	
Overall effect size		.09				.28		
80% credibility interval		–.07, .26				.02, .53		
T^2 / T^2_{res}		.02 / .01				.04 / .04		
Q-statistic (exact p value)		348.86* (.00)				856.08* (.00)		
I^2		71.31%				92.11%		

Note: *B* is the unstandardized regression coefficient, and β is the standardized regression coefficient. Values in parentheses following metaregression coefficients are the standard errors and the exact *p* values. R^2 refers to the percentage total variance explained, and R^2_{Meta} refers to the percentage of heterogeneity explained. T^2 is the between-studies variance, and T^2_{res} is the residual between-studies variance after accounting for moderators in the mixed-effects model. I^2 is the ratio of between-study variance to total variance. Power is expressed in terms the overall model and individual coefficients.
† $p < .10$.
* $p < .05$.

model and 1.00 per the ME model; however, the power values of the individual coefficients in the ME model were .67 for leading others, .84 for impact of decisions, and .57 for contact with others. We decided to proceed with evaluating the heterogeneity in our effects and with our metaregression analyses, with the important caveat that any one effect not reaching statistical significance in the equation predicting the extraversion–job satisfaction relation may be due to low power.

The second recommendation refers to the need to calculate several indices of heterogeneity. We performed the calculations in this step and all of the remaining steps using Viechtbauer's (2010) metafor package in R (syntax included in Appendix D), except for the 80% CrIs, which we calculated using the formulas of Schmidt and Hunter (2014). As shown in Table 6, for the relation between extraversion and job performance, the 80% CrI was [–.07, .26]; the T^2 statistic was .02; the Q statistic was 348.86 (p = .00); and the I^2 statistic was 71.31%. The wide CrI and the magnitude of the T^2 statistic, in conjunction with the statistically significant Q statistic, indicate sufficient heterogeneity to merit metaregression analyses. Furthermore, the I^2 statistic suggests that 71.31% of the variance in the effect sizes is due to between-studies variance (i.e., heterogeneity). As shown in Table 6, for the relation between extraversion and job satisfaction, the 80% CrI was [.02, .53]; the T^2 statistic was .04; the Q statistic was 856.08 (p = .00); and the I^2 statistic was 92.11%. The wide CrI and the magnitude of the T^2 statistic, in conjunction with the statistically significant Q statistic, indicate sufficient heterogeneity to merit metaregression analyses. Furthermore, the I^2 statistic suggests that 92.11% of the variance in the effect sizes is due to heterogeneity.

The third and very important recommendation is about choosing an appropriate model. As shown in Table 6, we present results via both the FE and ME models. We did this to illustrate the differences in conclusions across the two models; however, in this situation, the ME model is likely to be most appropriate given the large number of potential boundary conditions (which were not included in the model) of the extraversion–job performance and extraversion–job satisfaction relations.

In line with our fourth and fifth recommendations regarding the calculation of residual heterogeneity and testing boundary condition hypotheses, respectively, we used the restricted maximum likelihood estimator to calculate the residual heterogeneity (i.e., T^2_{res}) and the Knapp and Hartung (2003) method to test the statistical significance of the regression coefficients in the ME model. In the FE models, we used Hedges and Olkin's (2014) correction factor for the statistical significance tests of the regression coefficients.

Following our sixth recommendation regarding adjusting R^2, we used Aloe and colleagues' (2010) method to calculate R^2_{Meta}. As shown in Appendix D online, all of these functions are the default functions or are easily implemented with the metafor program.

Our next three recommendations are concerned with reporting results of a metaregression study. Following our seventh recommendation, we explained the rationale for all of our methodological choices in the preceding paragraphs. In line with our eighth recommendation, we reported both unstandardized and standardized coefficients. We used the unstandardized coefficients to calculate values of the relation between (a) extraversion and job performance and (b) extraversion and job satisfaction at different levels of the moderators. These results are presented as follows. In line with our ninth recommendation, we report the variable intercorrelations and descriptive statistics in Table 5.

Metaregression Results

With respect to the extraversion–job performance relation (Table 6), we found that impact of decisions (B = .05, β = .15, p = .01) and contact with others (B = .13, β = .32, p = .00) were statistically significant at the .05 level in the FE model, indicating moderate and large positive effects, respectively (Bosco et al., 2015). The three moderators explained 15% of the variance in the correlations between extraversion and job performance. Recall that because the FE model considers all variance to be attributable to sampling error, R^2 is equivalent to R^2_{Meta} in this case. In the ME model, only contact with others (B = .12, β = .24, p = .01) was statistically significant, indicating a large positive effect, and the model explained 12% of the heterogeneity in the correlations. In contrast, the R^2 indicated that the model explained 10% of the total variance in the correlations.

With respect to the extraversion–job satisfaction relation, we found that leading others (B = .09, β = .22, p = .00), impact of decision (B = –.04, β = –.12, p = .01), and contact with others (B = .18, β = .31, p = .01) were all statistically significant at the .05 level in the FE model. Thus, leading others and contact with others both had large positive effects, and impact of decisions had a small negative effect on the relation between extraversion and job satisfaction. The three moderators explained 15% of the variance in the correlations. In the ME model, none of the moderators were statistically significant at the .05 level, although it is important to note that this may be due to the low statistical power of the leading others and contact with others moderators. Furthermore, the model explained 4% of the heterogeneity in the effect sizes. In contrast, the uncorrected R^2 indicated that the model explained 6% of the total variance in the effect sizes. Although it is atypical for R^2_{Meta} to be less than R^2, it is possible under conditions where the moderators explain little of the heterogeneity in the effect sizes (Aloe et al., 2010).

As a final implementation of our recommendations regarding the interpretation of coefficients, we calculated values of the predicted correlations of extraversion with job performance and job satisfaction at different levels of the contact-with-others moderator. We used the ME results for these calculations. First, we calculated the predicted correlation between extraversion and job performance in jobs with a high vs. low degree of contact with others, operationalized as ±1 *SD* from the mean. We performed this calculation by substituting the values of contact with others that corresponded to ±1 *SD* (i.e., 4.99 and 4.31) and the values of the other moderators at their mean levels (i.e., 3.50 for leading others and 3.95 for impact) into the regression equation with unstandardized coefficients, as shown in Table 6. Although we do not show the intercept value of −.71 in the table, we used it to calculate these effect sizes. The result of this calculation indicated that, for jobs +1 *SD* in their level of contact with others, the predicted correlation between extraversion and job performance is .16, compared with .08 for jobs −1 *SD* in their level of contact with others. Second, we used the same procedure to calculate the predicted correlation between extraversion and job satisfaction in jobs with high contact with others. Because this was a different sample of studies, the mean values differed from those cited. Thus, the relevant value of contact with others at ±1 *SD* were 4.98 and 4.26; the mean value of leading others was 3.78; and the mean value of impact was 3.91. The intercept value in this equation was −.54. The result of this calculation indicated that, for jobs +1 *SD* in their level of contact with others, the predicted correlation between extraversion and job satisfaction is .30, as compared with .21 in jobs −1 *SD* in their level of contact with others.

Discussion

Metaregression is a powerful methodological tool for advancing theory because it uses all of the available data to examine a particular phenomenon, avoids the artificial dichotomization of moderator variables, and does not rely on the untenable assumption that reality includes only one boundary condition at a time. However, as evidenced by our literature review, researchers commonly make less-than-appropriate methodological choices that have implications for substantive conclusions, or they fail to report critical details that preclude readers from understanding what was done and precisely how conclusions were reached. Thus, the purpose of our article was to provide a single source that explains the various methodological issues and decision points that will influence substantive conclusions of metaregression studies.

Given the recent increase in the number of metaregression studies to test boundary conditions with population-level effects, we believe that it is critical for future studies to follow our best-practice recommendations to produce valid metaregression results.

Implications for Research and Practice

The issues that we identified in our review mean that in many cases the wrong words have been included in published articles' Results sections and the wrong words have been included in articles' Implications for Theory and Implications for Practice sections (cf. Cortina et al., 2017). To use our own illustrative metaregression studies, implementing the FE model would lead to the incorrect conclusion that two of the three moderators of the relation between extraversion and job performance were statistically significant and positive in direction such that the relation between extraversion and job performance becomes more strongly positive in jobs with high impact of decisions and high contact with others. Similarly, the FE model results showed that all three moderators were statistically significant in the model predicting the extraversion–job satisfaction relation (although, surprisingly, one was negative). However, the ME model revealed that only contact with others was a significant moderator of the extraversion–job performance relation and that none of the moderators were significant for the extraversion–job satisfaction relation. In addition, the FE model results indicated that the moderators explained more variance in the effect sizes (.15 for job performance, .15 for job satisfaction) than the ME results (.12 for job performance, .04 for job satisfaction). These disparate conclusions can have clear and tangible effects on research and practice, as the FE model results would support the propositions of theories of personality interactionism (e.g., Barrick et al., 2013) while suggesting that organizations interested in increasing job performance and satisfaction should invest resources in redesigning extraverts' jobs to involve more leadership opportunities and contact with other people. Furthermore, given differences between the FE and ME models in the magnitude of the effect in terms of the variance explained by the overall model and the magnitude of individual effect sizes, practitioners may falsely conclude that the benefits of a job redesign intervention will outweigh the costs when adopting the results of the FE model, when the ME model results may indicate that the benefits do not justify the costs of the intervention.

In light of our recommendations, we also suggest that part of the responsibility for advancing the usefulness and accuracy of metaregression analyses falls on reviewers and editors, in addition to authors. This is particularly the case given that reviewers and editors guide authors' decisions on the information that is presented in their papers. For example, Bosco et al. (2016) surveyed authors of 62 studies published in *Journal of Applied Psychology* and *Personnel Psychology* between 2005 and 2010 and found that 21% of respondents indicated that they had changed at least one hypothesis as a result of requests from the editor or reviewers. Thus, we suggest that reviewers and editors encourage authors conducting metaregression analyses to not only be as transparent as possible in reporting their methodological choices but also make appropriate decisions given their study context. In instances where journal space precludes the inclusion of an in-depth discussion of methodological choices, authors could include relevant information about methodological details in an online-only supplemental file, such as the results of a power analysis and results under different models (i.e., FE and ME).

Limitations and Future Research Directions

Although we sought to provide a comprehensive set of evidence-based recommendations on the use of metaregression analyses, there are other factors that we did not discuss that also influence meta-analytic conclusions more broadly and, by extension, conclusions from metaregression as well as other techniques that are based on meta-analytically derived effect sizes, such as meta-analytic structural equation modeling (Bergh et al., 2016). It is important to note that we did not discuss these factors because, to our knowledge, there is a lack of rigorous simulation or analytic work that demonstrates specific best practices. For example, duplicate and/or dependent effect sizes can influence meta-analytic conclusions, and various scholars have proposed different ways to detect and handle them (J. A. Wood, 2008). The most commonly used method for dealing with dependent effect sizes, which we also used in our illustrative metaregression studies, is to compute the average of multiple effect sizes from individual studies, which results in a single effect size estimate per study (Geyskens et al., 2009). We made this choice to simplify our analysis and focus on the implementation of our best-practice recommendations. However, this approach has been criticized as being less accurate than other approaches, such as calculating composites (Schmidt & Hunter, 2014) and adjusting standard errors based on the degree of dependency (Beaty et al., 2011). However, little research has examined how dependent effect sizes affect metaregression specifically as well as how to incorporate corrections for dependency into ME models. Certainly, this is an issue that deserves close attention in future research.

Second, we did not conduct outlier analyses in our sample metaregression and do not offer any specific recommendations regarding outliers in the context of metaregression. Geyskens et al. (2009) offered a summary of available approaches to dealing with outliers in meta-analyses; however, given that metaregression also incorporates assumptions from regression, there is a need to integrate Geyskens and colleagues' (2009) recommendations with methods meant to deal with outliers more generally in a regression context (e.g., Aguinis, Gottfredson, & Joo, 2013). This is important given our findings that half of metaregression studies included in our review relied on ≤37 studies and such small sample sizes are particularly liable to be affected by outliers. Therefore, future research should examine how to handle outliers in a metaregression context.

Conclusion

We emphasize that metaregression is not a panacea, as there are several instances when metaregression analyses will not be useful in terms of producing meaningful theory advancements—namely, when there is insufficient power to detect boundary conditions or when there is little heterogeneity in the population of primary-level studies. However, metaregression is a potentially powerful tool for advancing theory. We hope that our evidence-based recommendations will serve as a catalyst for using metaregression to produce knowledge that will advance theory and improve our understanding of conditions under which practical applications are more or less likely to be successful.

Notes

1. Although a 5-point scale is not perfectly continuous, it provides a significant improvement over a 2-point scale (Aguinis, Pierce, & Culpepper, 2009) in terms of the accuracy of substantive results. For example, we used the online calculator described in Aguinis et al. (2009) and found that an observed correlation of .25 obtained with 2-point scales for the predictor and criterion corresponds to an underlying true correlation of .38 (i.e., a downward bias of .13 correlation points). This bias is much smaller if 5-point scales are used: .03 correlation points.

2. An ME model is analogous to a random-effects model in general meta-analysis, with the difference being that an ME model also incorporates a fixed factor into the equation (i.e., the boundary condition).

3. As noted by an anonymous reviewer, it is possible that in many cases reviewers or editors may have asked authors to omit information about power analysis to save journal space.

4. These include handling duplicate and/or dependent effect sizes, identifying and handling outliers, reliably coding study data and moderators, and correcting for study artifacts, among others. We refer to recommendations offered by Aguinis, Dalton, Bosco, Pierce, and Dalton (2010); Aytug, Rothstein, Zhou, and Kern (2012); Geyskens et al. (2009); Kepes, McDaniel, Brannick, and Banks (2013); and J. A. Wood (2008).

5. The SAS macro is available from https://static-content.springer.com/esm/art%3A10.3758%2FBRM.41.1.35/MediaObjects/Cafri-BRM-2009.zip.

References

Aguinis, H. 1995. Statistical power with moderated multiple regression in management research. *Journal of Management*, 21: 1141-1158.

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. A., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. 2010. Meta-analytic choices and judgement calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, 37: 5-38.

Aguinis, H., Forcum, L. E., & Joo, H. 2013. Using market basket analysis in management research. *Journal of Management*, 39: 1799-1824.

Aguinis, H., & Glavas, A. 2017. On corporate social responsibility, sensemaking, and the search for meaningfulness through work. *Journal of Management*. Advance online publication. doi:10.1177/0149206317691575

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. 2013. Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, 39: 1490-1528.

Aguinis, H., Gottfredson, R. K., & Joo, H. 2013. Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16: 270-301.

Aguinis, H., Gottfredson, R. K., & Wright, T. A. 2011. Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior*, 32: 1033-1043.

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. The 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., & Stone-Romero, E. F. 1997. Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, 82: 192-206.

Aloe, A. M., Becker, B. J., & Pigott, T. D. 2010. An alternative to R^2 for assessing linear models of effect size. *Research Synthesis Methods*, 1: 272-283.

American Psychological Association. 2008. Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63: 839-851.

Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. 2012. Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. *Organizational Research Methods*, 15: 103-133.

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

Barrick, M. R., Mount, M. K., & Li, N. 2013. The theory of purposeful work behavior: The role of personality, higher-order goals, and job characteristics. *Academy of Management Review*, 38: 132-153.

Beaty, J. C., Nye, C. D., Borneman, M. K., Kantrowitz, T. M., Drasgow, F., & Grauer, E. 2011. Proctored versus unproctored internet tests: Are unproctored noncognitive tests as predictive of job performance? *International Journal of Selection and Assessment*, 19: 1-10.

Bedeian, A. G. 2013. "More than meets the eye": A guide to interpreting the descriptive statistics and correlation matrices reported in management research. *Academy of Management Learning & Education*, 12: 121-135.

Bergh, D. D., Aguinis, H., Heavy, C., Ketchen, D. J., Boyd, B. K., Su, P., Lau, C., & Joo, H. 2016. Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal*, 37: 477-497.

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. 2017. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization*. Advance online publication. doi:10.1177/1476127017701076

Bernerth, J. B., & Aguinis, H. 2016. A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, 69: 229-283.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. 2010. Transfer of training: A meta-analytic review. *Journal of Management*, 36: 1065-1105.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. 2009. *Introduction to meta-analysis*. Chichester, UK: Wiley.

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. 2016. Harking's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology*, 69: 709-750.

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.

Boyd, B. K., Haynes, K. T., Hitt, M. A., Bergh, D. D., & Ketchen, D. J. 2012. Contingency hypotheses in strategic management research: Use, disuse, or misuse? *Journal of Management*, 38: 278-313.

Busse, C., Kach, A. P., & Wagner, S. M. 2016. Boundary conditions: What they are, how to explore them, why we need them, and when to consider them. *Organizational Research Methods*. Advance online publication. doi:10.1177/1094428116641191

Cafri, G., Kromrey, J. D., & Brannick, M. T. 2009. A SAS macro for statistical power calculations in meta-analysis. *Behavior Research Methods*, 41: 35-46.

Carney, M., Gedajlovic, E. R., Heugens, P. M. A. R., van Essen, M., & van Oosterhout, J. H. 2011. Business group affiliation, performance, context, and strategy: A meta-analysis. *Academy of Management Journal*, 54: 437-460.

Cohen, J. 1992. Statistical power analysis. *Current Directions in Psychological Science*, 1: 98-101.

Cortina, J. M., Aguinis, H., & DeShon, R. P. 2017. Twilight of dawn or of evening? A century of research methods in the *Journal of Applied Psychology*. *Journal of Applied Psychology*, 102: 274-290.

D'Innocenzo, L., Mathieu, J. E., & Kukenberger, M. R. 2014. A meta-analysis of different forms of shared leadership–team performance relations. *Journal of Management*, 42: 1964-1991.

Drees, J. M., & Heugens, P. M. A. R. 2013. Synthesizing and extending resource dependence theory: A meta-analysis. *Journal of Management*, 39: 1666-1698.

Erez, A., Bloom, M. C., & Wells, M. T. 1996. Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology*, 49: 275-306.

Fiedler, F. E. 1967. *A theory of leadership effectiveness*. New York: McGraw-Hill.

Geyskens, I., Krishnan, R., Steenkamp, J. B. E., & Cunha, P. V. 2009. A review and evaluation of meta-analysis practices in management research. *Journal of Management*, 35: 393-419.

Gonzalez-Mulé, E. 2015. *Contextual job features and occupational values as moderators of personality trait validities: A test and extension of the theory of purposeful work behavior*. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Gonzalez-Mulé, E., Mount, M. K., & Oh, I. S. 2014. A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology*, 99: 1222-1243.

Griffeth, R. W., Hom, P. W., & Gaertner, S. 2000. A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26: 463-488.

Gurevitch, J., & Hedges, L. V. 1999. Statistical issues in ecological meta-analyses. *Ecology*, 80: 1142-1149.

Hackman, J. R., & Oldham, G. R. 1975. Development of the job diagnostic survey. *Journal of Applied Psychology*, 60: 159-170.

Hall, J. A., & Rosenthal, R. 1991. Testing for moderator variables in meta-analysis: Issues and methods. *Communications Monographs*, 58: 437-448.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. 2007. Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92: 373-385.

Heavey, A. L., Holwerda, J. A., & Hausknecht, J. P. 2013. Causes and consequences of collective turnover: A meta-analytic review. *Journal of Applied Psychology*, 98: 412-453.

Hedges, L. V., & Olkin, I. 1985. *Statistical methods for meta-analysis*. New York: Academic Press.

Hedges, L. V., & Olkin, I. 2014. *Statistical methods for meta-analysis*. New York: Academic Press.

Hedges, L. V., & Pigott, T. D. 2004. The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9: 426-445.

Heugens, P. M. A. R., & Lander, M. W. 2009. Structure! Agency! (And other quarrels): A meta-analysis of institutional theories of organization. *Academy of Management Journal*, 52: 61-85.

Higgins, J., & Thompson, S. G. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21: 1539-1558.

Higgins, J., & Thompson, S. G. 2004. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, 23: 1663-1682.

Hüffmeier, J., Freund, P. A., Zerres, A., Backhaus, K., & Hertel, G. 2014. Being tough or being nice? A meta-analysis on the impact of hard- and softline strategies in distributive negotiations. *Journal of Management*, 40: 866-892.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. 1982. *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Judge, T. A., & Zapata, C. P. 2015. The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal*, 58: 1149-1179.

Kepes, S., & McDaniel, M. A. 2013. How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology*, 6: 252-268.

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. 2013. Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-Analytic Reporting Standards). *Journal of Business and Psychology*, 28: 123-143.

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.

Knapp, G., & Hartung, J. 2003. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22: 2693-2710.

Locke, E. A., & Latham, G. P. 1990. *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.

López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W., & Viechtbauer, W. 2014. Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *British Journal of Mathematical and Statistical Psychology*, 67: 30-48.

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. 2012. Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, 97: 951-966.

O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. 2017. The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*, 43: 376-399.

Overton, R. C. 1998. A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3: 354-379.

Park, T. Y., & Shaw, J. D. 2013. Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology*, 98: 268-309.

Ployhart, R. E., & Vandenberg, R. J. 2010. Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, 36: 94-120.

Pocock, S. J., Cook, D. G., & Beresford, S. A. 1981. Regression of area mortality rates on explanatory variables: What weighting is appropriate? *Journal of the Royal Statistical Society*, 30: 286-295.

Rabl, T., Jayasinghe, M., Gerhart, B., & Kühlmann, T. M. 2014. A meta-analysis of country differences in the high-performance work system–business performance relationship: The roles of national culture and managerial discretion. *Journal of Applied Psychology*, 99: 1011-1041.

Raudenbush, S. W. 2009. Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 296-314). New York: Russell Sage Foundation.

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. 2009. Measuring organizational performance: Towards methodological best practice. *Journal of Management*, 35: 718-804.

Schmidt, F. L., & Hunter, J. E. 1999. Comparison of three meta-analysis methods revisited: An analysis of Johnson, Mullen, and Salas (1995). *Journal of Applied Psychology*, 84: 144-148.

Schmidt, F. L., & Hunter, J. 2004. General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86: 162-173.

Schmidt, F. L., & Hunter, J. E. 2014. *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Sidik, K., & Jonkman, J. N. 2007. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine*, 26: 1964-1981.

Steel, P. D., & Kammeyer-Mueller, J. D. 2002. Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology*, 87: 96-111.

Steel, P., Kammeyer-Mueller, J., & Paterson, T. A. 2015. Improving the meta-analytic assessment of effect size variance with an informed Bayesian prior. *Journal of Management*, 41: 718-743.

Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. 2015. The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research. *Journal of Management*, 41: 491-520.

Sturman, M. C. 2003. Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management*, 29: 609-640.

Sturman, M. C., Cheramie, R. A., & Cashen, L. H. 2005. The impact of job complexity and performance measurement on the temporal consistency, stability, and test-retest reliability of employee job performance ratings. *Journal of Applied Psychology*, 90: 269-283.

Thompson, S. G., & Higgins, J. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21: 1559-1573.

Thompson, S. G., & Sharp, S. J. 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, 18: 2693-2708.

Van den Noortgate, W., & Onghena, P. 2003. Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, 35: 1-10.

van Essen, M., Otten, J., & Carberry, E. J. 2015. Assessing managerial power theory: A meta-analytic approach to understanding the determinants of CEO compensation. *Journal of Management*, 41: 164-202.

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. 2017. A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management*. Advance online publication. doi:10.1177/0149206317702220

Viechtbauer, W. 2010. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36: 1-48.

Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. 2015. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods*, 20: 360-374.

Webber, S. S., & Donahue, L. M. 2001. Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. *Journal of Management*, 27: 141-162.

Whitener, E. M. 1990. Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, 75: 315-321.

Wilson, D. B. 2005. *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved from http://mason.gmu.edu/~dwilson/ma.html.

Wood, J. A. 2008. Methodology for dealing with duplicate study effects in a meta-analysis. *Organizational Research Methods*, 11: 79-95.

Wood, R. E., Mento, A. J., & Locke, E. A. 1987. Task complexity as a moderator of goal effects: A meta-analysis. *Journal of Applied Psychology*, 72: 416-425.

Wright, P. M. 2016. Ensuring research integrity: An editor's perspective. *Journal of Management*, 42: 1037-1043.

Zhong, W., Su, C., Peng, J., & Yang, Z. 2014. Trust in interorganizational relationships: A meta-analysis integration. *Journal of Management*. Advance online publication. doi:10.1177/0149206314546373

Zyphur, M. J., & Oswald, F. L. 2015. Bayesian estimation and inference: A user's guide. *Journal of Management*, 41: 390-420.