Times are changing, Bias isn't

Siegel, Magdalena; Eder, Junia Sophia Nur; Wicherts, Jelte M.; Pietschnig, Jakob

Published in:
Journal of Applied Psychology

DOI:
10.1037/apl0000991

Publication date:
2022

Document Version
Peer reviewed version

Times are Changing, Bias isn't: A Meta-Meta-Analysis on Publication Bias Detection Practices, Prevalence Rates, and Predictors in Industrial/Organizational Psychology

Magdalena Siegel[1,2], Junia Sophia Nur Eder[1], Jelte M. Wicherts[3], & Jakob Pietschnig[1]

[1]Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Austria

[2]Department of Public Health, Institute of Tropical Medicine Antwerp, Belgium

[3]Department of Methodology and Statistics, Tilburg University, Netherlands

Siegel, M., Eder, J. S. N., Wicherts, J. M., & Pietschnig, J. (2021). Times are changing, bias isn't: A meta-meta-analysis on publication bias detection practices, prevalence rates, and predictors in industrial/organizational psychology. *Journal of Applied Psychology*. Advance online publication. https://doi.org/10.1037/apl0000991

Author Note

Data and R code to reproduce analyses and figures are provided under https://osf.io/dqc3y/.

The research by MS, JSNE, and JP was funded by the Austrian Science Fund (FWF): project number P31002. The research by JMW was supported by a Consolidator Grant 726361 (IMPROVE) from the European Research Council (ERC, https://erc.europa.eu).

We would like to thank all authors who generously provided us with their data or additional information to their meta-analyses. We would also like to thank Jana Kühnel and Ingrid Wahl for providing their expertise on taxonomies within the field of I/O Psychology.

Correspondence concerning this article should be addressed to Magdalena Siegel, Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria. E-Mail: magdalena.siegel@univie.ac.at

Abstract

Effect misestimations plague Psychological Science, but advances in the identification of dissemination biases in general and publication bias in particular have helped in dealing with biased effects in the literature. However, the application of publication bias detection methods appears to be not equally prevalent across subdisciplines. It has been suggested that particularly in I/O Psychology, appropriate publication bias detection methods are underused. In this meta-meta-analysis, we present prevalence estimates, predictors, and time trends of publication bias in 128 meta-analyses that were published in the *Journal of Applied Psychology* (7,263 effect sizes, 3,000,000+ participants). Moreover, we reanalyzed data of 87 meta-analyses and applied nine standard and more modern publication bias detection methods. We show that (i) the bias detection method applications are underused (only 41% of meta-analyses use at least one method) but have increased in recent years, (ii) those meta-analyses that apply such methods now use more, but mostly inappropriate methods, and (iii) the prevalence of potential publication bias is concerning but mostly remains undetected. Although our results indicate somewhat of a trend towards higher bias awareness, they substantiate concerns about potential publication bias in I/O Psychology, warranting increased researcher awareness about appropriate and state-of-the-art bias detection and triangulation. Embracing open science practices such as data sharing or study preregistration is needed to raise reproducibility and ultimately strengthen Psychological Science in general and I/O Psychology in particular.

Keywords: dissemination bias, publication bias, meta-meta-analysis, industrial and organizational psychology, decline effect

Times are Changing, Bias isn't: A Meta-Meta-Analysis on Publication Bias Detection Practices, Prevalence Rates, and Predictors in Industrial/Organizational Psychology

For decades, meta-analyses have contributed substantially to the advancement of the field of Industrial and Organizational (I/O) Psychology (Rothstein et al., 2005; Schmidt & Hunter, 2015). However, concerns about publication bias in this field (e.g., Kepes et al., 2012) have regained attention in the course of the recent confidence crisis (Ioannidis, 2005b; Open Science Collaboration, 2015). Publication bias is a well-known dissemination bias (see Fanelli et al., 2017; Rothstein et al., 2005, for an overview) and occurs when published research on a given topic is unrepresentative of all the research that has been carried out on this topic (Rothstein et al., 2005). Typically, publication bias manifests itself in inflated effect sizes (although effect deflation has also been reported in the literature in certain cases, e.g., McDaniel et al., 2006) because non-significant studies (or analyses) with correspondingly small effects are being published less frequently (i.e., they are withheld from publication; Rothstein et al., 2005).

This poses a threat to the credibility of scientific findings, because inflated effect sizes serve as statistical and also conceptual "false anchors" by biasing subsequent power-calculations and serving as authoritative sources due to their higher citation rates (Greenberg, 2009). On the meta-analytic level, publication bias has been shown to negatively affect moderation analyses and heterogeneity assessments, often in unpredictable ways (Augusteijn et al., 2019). The detrimental effect of publication bias is exacerbated in meta-analyses, as research syntheses are of higher evidential value than primary studies, are cited more frequently (Aguinis et al., 2011), and are designed to inform evidence-based policy and practice.

Publication Bias: Reasons and Ramifications

Reasons for publication bias are manifold and rooted and perpetuated in the current scientific reward system that favors the publication of novel, significant, and hypothesisconforming results over the publication of replication studies, non-significant results, or hypothesis non-conforming results (Fanelli, 2012; Ioannidis, 2005a, 2005b). In such a system, studies (or analyses) yielding non-significant results are rejected during the submission process or do not even enter the submission stage (Cooper et al., 1997), because authors see no merit in the manuscript preparation. Moreover, authors under high pressure to publish may engage in questionable research practices (QRPs, see Wicherts et al., 2016) to obtain statistically significant results that are in line with their hypotheses, or may formulate their hypotheses post-hoc and in line with their significant findings (a practice known as HARKing, meaning hypothesizing after results are known; Kerr, 1998). Direct evidence for selective reporting of significant results has been amassed in many scientific fields (van der Steen et al., 2018; see also Wicherts, 2017, for an overview) and evidence accumulates that I/O Psychology is no exception to fostering these behaviors (see Banks et al., 2016, for a review; Bedeian et al., 2010; O'Boyle et al., 2019).

Publication Bias in I/O Psychology

Repeated concerns have been voiced about I/O Psychology's reluctance in acknowledging the importance of publication bias and its consequences by some researchers (Banks & McDaniel, 2011; Kepes et al., 2012; but see Dalton et al., 2012, for a differing view, and Schmidt & Hunter, 2015, for an in-depth account). These concerns are based on comparatively low prevalence rates of studies identifying bias in this field, ranging between 18% (Aytug et al., 2012) and 31% (Banks et al., 2012; Kepes et al., 2012). Those meta-analyses that do report results from bias analyses typically only use comparatively few, outdated (e.g., the fail-safe *N*; Rosenthal, 1979), or otherwise inappropriate (e.g., the direct comparison between published and unpublished effect sizes; see Aytug et al., 2012; Kepes et al., 2012) publication bias detection methods (Kepes et al., 2012; O'Boyle et al., 2014).

This adds to the concern of under-detection of publication bias in I/O Psychology, because many meta-analysts who attempt to investigate bias may fail to detect it due to their methodological choices. Moreover, I/O Psychology has traditionally favored an approach to meta-analysis that strives to correct for potential effect deflation due to statistical artifacts (e.g., unreliability) rather than effect inflation (i.e., psychometric meta-analysis; Schmidt & Hunter, 2015). Because the available publication bias detection methods fail to correct for statistical artifacts (Schmidt & Hunter, 2015), applications of these detection tools may be less common in I/O Psychology.

Recent investigations have shown that I/O Psychology and related disciplines are not immune to publication bias. Reanalyses of meta-analyses found at least some indication of publication bias in 12 out of 15 (entrepreneurship research; O'Boyle et al., 2014) and three out of four examined cases (information systems; Kepes & Thomas, 2018). Similarly, effect sizes in strategic management research have been shown to be inflated by about 30% (Harrison et al., 2017), echoing prior concerns in this field (Miller & Tsang, 2011). Publication bias has also been shown to affect specific research topics, such as test validities (McDaniel et al., 2006), stereotype threat (Nguyen & Ryan, 2008; Shewach et al., 2019; Zigerell, 2017), personality correlates of job performance (Kepes & McDaniel, 2015), or employee turnover (Field et al., 2020). This bears real-world implications because managerial and organizational decisions and practices are based on distorted evidence (Field et al., 2020; Kepes et al., 2012).

It is important to note that evidence for potential publication bias in the field of I/O Psychology has not been found to be ubiquitous or equally prevalent across subdisciplines (Kepes et al., 2014). For example, in one of the earliest investigations into the "file-drawer problem", no evidence of publication bias (here defined as evidence of unpublished studies in the "file drawers" of researchers) was found (Campbell, 1990). Similarly, no publication bias (defined as a larger share of significant published than unpublished correlations) was found in a large-scale examination of correlations used as input for correlation matrices and meta-analyses (Dalton et al., 2011; see also Dalton et al., 2012). In addition, findings suggest that publication bias might be of little relevance in fields with two or more competing theories (rather than a single dominant theory, Doucouliagos & Stanley, 2013), or when effect sizes that were meta-analyzed were of incidental nature in primary studies (rather than the main hypothesis, Schmidt & Hunter, 2015, Schmidt et al., 2009). The conflicting findings outlined above – coupled with the notion that publication bias detection methods are based on different methodological families (and thus, different definitions of publication bias), are not without limitations (see below and Supplement S1), and do not incorporate artifact correction (Schmidt & Hunter, 2015) – raise the question about the conceptualization, degree, and severity of publication bias in the field.

Low publication bias detection prevalence coupled with empirical findings on the existence of publication bias in several subfields might indicate under-detection of publication bias. However, the recent replicability crisis (Open Science Collaboration, 2015) sparked an upsurge in reporting guidelines for meta-analyses (American Psychological Association, 2010; Moher et al., 2009; Schalken & Rietbergen, 2017), that advise researchers to investigate bias-related mechanisms in their syntheses. Thus, the intensifying replicability debate in the field (Landis & Rogelberg, 2013) and methodological advances in empirical bias detection (Stanley & Doucouliagos, 2014; van Aert et al., 2016) may have enhanced the awareness in applied meta-analysts about publication bias in I/O Psychology.

Without doubt, meta-analysts from previous decades can certainly neither be blamed for using methods that were deemed appropriate then (but are now considered to be outdated) nor for not using methods that had yet to be developed. But when keeping in mind the upsurge in the development of novel bias detection methods in the past years, it remains unclear if that is reflected in the application practices of meta-analysts.

Methods to Detect Publication Bias

The advent of meta-analysis (Glass, 1976) brought along empirical methods to detect publication bias, the first being the failsafe-*N* (Rosenthal, 1979). In the following decades, several methodological families aiming at investigating different aspects of publication bias emerged. For one, methods designed to investigate so-called small study effects (i.e., studies with smaller sample sizes are systematically different [typically: larger] than effect sizes from studies with larger sample sizes; Sterne et al., 2005) were developed. These include Begg and Mazumdar's rank test (Begg & Mazumdar, 1994), Sterne and Egger's regression (Egger et al., 1997; Sterne & Egger, 2005), the trim-and-fill approach (Duval & Tweedie, 2000a, 2000b), PET-PEESE (Stanley & Doucouliagos, 2014), as well as graphical displays such as funnel plot variants (Light & Pillemer, 1984; Peters et al., 2008) or cumulative meta-analyses sorted by study precision (Kepes et al., 2012).

Second, methods based on *p*-values have been developed in the 2010s, namely *p*-curve (Simonsohn et al., 2014, 2015), *p*-uniform (van Aert et al., 2016; van Assen et al., 2015), and its recent extension *p*-uniform* (van Aert & van Assen, 2018). Third, the test of excess significance (Ioannidis & Trikalinos, 2007), which compares the number of expected and observed significant effect sizes, has been developed. Fourth, selection model approaches (initially introduced in the 1980s, e.g., Iyengar & Greenhouse, 1988) have regained popularity, both due to the development of variants that are suitable for smaller datasets (Vevea & Woods, 2005), as well as due to their favorable performance in simulation studies (McShane et al., 2016). The detailed characteristics of these methods are provided in the Supplement S1.

Currently, there is no single method that would have been shown to clearly outperform all others (Carter et al., 2019; Renkewitz & Keiner, 2019). Instead, it has been observed that different methods do not work equally well under different conditions. Methods differ with regard to their susceptibility to false-positive rates under between-studies heterogeneity (Peters et al., 2006), their statistical power to detect publication bias (Ioannidis & Trikalinos, 2007), their accuracy in effect size estimation when *p*-hacking is present (van Aert et al., 2016), or the number of effect sizes in a meta-analytic dataset (Sterne & Egger, 2005).

Furthermore, these methods were designed under different conceptual assumptions and with different aims: Whereas some methods aim to detect small study effects (e.g., Trim-and-Fill), others aim to provide an indication of the evidential value of a set of effect sizes (e.g., *p*-curve), while others (such as selection model approaches) aim to provide an adjusted estimate under differing scenarios of bias severity. Thus, current recommendations favor a broadband-approach (i.e., bias triangulation) that necessitates including several publication bias analyses within a given meta-analysis to account for the differing strengths and weaknesses of these methods (Carter et al., 2019; Schmidt & Hunter, 2015, Vevea et al., 2019) as well as to capture different operationalizations of publication bias. Importantly and in light of inherent shortcomings of current bias detection methods, these analyses should be seen as sensitivity analyses (Schmidt & Hunter, 2015, Vevea & Woods, 2005), rather than as definitive indications for publication bias.

Predictors of Publication Bias

Because publication bias is assessed at the meta-analytic level, predictors of publication bias can be assessed at the meta-meta-analytic level. On the primary study level, the (relative) effect strength of individual studies may be related to the presence of publication bias in a meta-analysis. This assumption is based on mechanisms that are responsible for the so-called decline effect (Schooler, 2011). The decline effect describes the phenomenon that effect sizes tend to decrease in strength over time, regardless of the addressed research question (see Protzko & Schooler, 2017). Indeed, declining effects seem to be a common phenomenon in Psychology (e.g., Pietschnig et al., 2010, 2015, 2019).

Declining effects may be rooted in the publication of a particularly large, significant, but inflated (or outright false) effect from an exploratory, underpowered study (e.g., a test of a novel treatment in a small sample that reaches significance by chance). Subsequently, this (inflated) effect may affect power-estimations of future replication attempts and prompt researchers to either not publish non-significant findings (because they are based on samples that are too small to detect a true non-zero effect) or engage in QRPs to obtain a significant (but inflated) and publishable result (Pietschnig et al., 2019). Thus, meta-analyses showing declining effects may also be expected to show more indications of publication bias.

On the meta-analytic level, we propose that the strength of the meta-analytic summary effect (when taken as a proxy for the underlying true effect size) is negatively related to publication bias indication. This assumption is similarly rooted in power-considerations and subsequent submission behavior by authors: Large true effects can be detected (i.e., reach nominal significance) even in small samples with low power, thus not necessitating QRPs or strategic submission behaviors.

Considering the known pitfalls of current publication bias detection methods, it is important to distinguish genuine predictors of publication bias from non-genuine ones (e.g., between study heterogeneity has been found to produce false positive rates in some methods; Peters et al., 2006, some methods suffer from low power when the number of included effect sizes is small, Sterne et al., 2005). Therefore, publication bias indication may also be related to between-study heterogeneity and number of effect sizes within a meta-analysis.

The Present Meta-Meta-Analysis

The present meta-meta-analysis has two aims: First, we investigate prevalence and time trends of publication bias detection methods and author-reported bias rates in meta-analyses published in the *Journal of Applied Psychology.* This journal is a flagship journal in I/O Psychology with a longstanding tradition of frequently publishing meta-analyses of high quality and reporting transparency (Aytug et al., 2012). Second, we reanalyze meta-analytic datasets reported therein using nine standard and more modern publication bias detection methods and several operationalizations of bias as documented in the literature. By means of hierarchical logistic and Poisson meta-meta-regression models, we additionally assess meta-analytic and study-level predictors of potential bias indication.

We extend previous works in four important ways: First, previous investigations in the field of I/O Psychology as a whole (e.g., Aytug et al., 2012; Kepes et al., 2012) predated both the development of more modern approaches to assess publication bias as well as a general shift towards open science practices in I/O Psychology. Moreover, changes in detection practices have yet to be comprehensively addressed because of the limited timeframe of earlier investigations. Second, more recent investigations into publication bias-related mechanisms in I/O Psychology and related disciplines (e.g., Field et al., 2020; Harrison et al., 2017; O'Boyle et al., 2014) focused on rather narrow aspects of certain subfields or topics and thus may not be reflective of patterns that characterize a subdiscipline such as I/O Psychology on a broader scale. Third, by applying a broad range of publication bias detection methods with different rationales and thresholds to various combinations and stratifications of meta-analytic datasets gathered from the field of I/O Psychology, we provide a comprehensive yet in-depth and multi-faceted analysis of potential bias indication. Fourth, the current study is one of the first investigations in (I/O) Psychology that directly link meta-analytic and primary study characteristics to publication bias indication using two different meta-meta-regression designs. In all, our meta-meta-analysis provides an up-to-date (through September 2020) and to our knowledge the most comprehensive assessment of trends in and drivers of publication bias detection and potential bias indication within the field of I/O Psychology to date.

Method

Information Sources

The present meta-meta-analysis is part of a larger project that aims at investigating publication and other dissemination biases in the field of Psychological Science (see https://osf.io/8w5zc and https://osf.io/s8z6y for project preregistrations). For the present study, we identified meta-analyses published in the *Journal of Applied Psychology* by searching the online database ISI Web of Knowledge (date of search: January 5, 2018; update:

September 15, 2020) using the search string ‘meta-analy* OR "research synthes*"’ (topic) AND ‘journal of applied psychology’ (publication name). We deemed this to be a reasonable approach, because all volumes of the *Journal of Applied Psychology* are indexed in ISI Web of Knowledge.

Inclusion and Exclusion Criteria

Meta-analyses were deemed eligible for inclusion if they fulfilled the following six criteria: First, they had to report an original meta-analysis by statistically synthesizing effect sizes extracted from primary published or unpublished studies. Thus, comments or primary studies, meta-analytic simulation studies (e.g., Hattrup et al., 1997), meta-analytic methodology papers (e.g., Cheung & Chan, 2004), or reanalyses of previously published meta-analyses (e.g., Zigerell, 2017) were excluded.

Second, meta-analyses had to be carried out using traditional meta-analytic approaches as outlined by Hedges and Olkin (1985) or Hunter and Schmidt (e.g., Hunter & Schmidt, 2004). Bayesian meta-analyses, network meta-analyses, or other non-standard meta-analyses were not eligible for inclusion. Studies employing meta-analytic structural equation models were included only if they provided the bivariate meta-analytic correlation matrices that were used as input for the structural equation modeling.

Third, the meta-analyses had to be based on traditional effect sizes for bivariate associations or group differences, namely, Pearson r, Fisher z, Cohen d, Hedges g, Odds Ratios, or Log Odds Ratios. Meta-analyses based on risk ratios or hazard ratios were eligible for inclusion if the necessary cell counts were provided to calculate Fisher z values (see section ‘Preparation of Primary Data’). Meta-analyses using other, non-standard effect sizes such as interrater reliabilities (e.g., Conway et al., 1995) were excluded.

Fourth, the timeframe for the literature search within a given meta-analysis should not have been arbitrarily restricted. This means that a conceptual justification (e.g., the development of a new assessment method or publication of theory) had to be provided within the meta-analysis if the time period of the literature search was in any way restricted (e.g., only the past ten years). This inclusion criterion was used because it is likely that the severity of publication biases in any given field changes over time, which may mask biases, particularly the decline effect investigated in this study, in arbitrarily time-restricted meta-analyses. For example, we excluded a meta-analysis that used a starting date for their literature search six years after the publication of a landmark study that investigated the relevant research question (Kinicki et al., 2002) which could have led to the exclusion of several relevant studies.

Fifth, the meta-analysis had to provide enough information to recalculate the meta-analytic summary effect reported therein. This means that for each primary study included in the meta-analysis, either (i) effect sizes and sample sizes (or standard errors) or (ii) test statistics that could be transformed to Fisher *z* had to be reported. If this information had not been provided in text or in supplementary files, we requested the data by contacting the corresponding authors of the respective meta-analysis. We sent a first reminder after two weeks and a second reminder after four weeks after our initial request in cases where we did not receive any reply from the authors. If we did not obtain a reply after the second reminder or if the authors of the meta-analyses indicated unavailability of the data, the meta-analyses were excluded from primary data reanalyses.

Sixth, if the full primary data were unobtainable from a meta-analysis, the effect size and sample size of the initial study (i.e., the oldest study in terms of publication years) had to be identifiable beyond reasonable doubt from the initial study itself. This inclusion criterion was specified in line with the aims of the larger context of the overall project (https://osf.io/8w5zc) and because the strength of the initial effect was included as a predictor in our meta-meta-regression models. In cases where study authors were not able to provide us with the full dataset (e.g., for meta-analyses published before or in the early 2000s), we additionally inquired for the provision of the initial study effect. This approach enabled us to include 31 meta-analyses (14 additional datasets and 17 initial effects out of 124 data requests; including request for several studies to the same authors).

Of note, primary data availability (either reported in the paper or provided by authors) did not significantly predict effect strength of the reported meta-analytic summary effect, $t(66.26) = -1.11$, $p = .269$, the likelihood of investigating publication bias ($b = 0.58$, $p = .188$), or the number of methods used ($b = -0.15$, $p = .556$; both models controlling for differences in data availability due to publication year). Still, we cannot rule out a slight bias towards studies with higher reporting quality, thus our estimate of publication bias prevalence may represent an upper threshold of the actual use of these methods in I/O Psychology.

Study Selection and Data Extraction

We assessed full texts of all 362 hits retrieved by our literature search. In line with best-practice recommendations, each study was coded twice: Two authors [MS, JP] independently coded about two thirds (251 or 69.34%) of studies, and the remaining third was double-coded by one author [MS]. Discrepancies were resolved by discussion between the two authors ($Mdn_{\%} = 86.89\%$, $range_{\%} = 73.77\% – 100\%$; Mdn_{κ} for categorical variables = .71, $range_{\kappa} = .58 – 1.00$).

To ensure independence on a meta-meta-analytic level, we only coded one individual meta-analysis in each study, if more than one meta-analysis had been reported. The respective meta-analysis was selected according to the following criteria: We first identified the meta-analysis (or meta-analyses) that the initial study contributed to. If this study contributed to more than one individual meta-analysis (e.g., it provided effect sizes for the association between several personality dimensions and job performance, each of which had been meta-analyzed separately), we included the meta-analysis comprising the largest number of effect sizes (for a similar approach, see Ferguson & Brannick, 2012). We favored this selection criterion over other criteria (e.g., random selection of a meta-analysis that the oldest study contributed to) to ensure an adequate meta-analytic sample size, which is beneficial for many publication bias detection methods (Cooper et al., 2019). If the oldest study contributed to several meta-analyses with an identical numbers of effect sizes, we randomly selected one for inclusion.

Of note, this decision criterion implied that we always selected (i) the main meta-analysis (instead of subgroup analyses) that the oldest study contributed to, as well as (ii) meta-analyses including outliers (instead of sensitivity analyses without outliers that may have been conducted by the study authors). However, we do not assume that this systematically biased our meta-meta-analytic sample towards particularly large or heterogenous meta-analyses, since our first selection criterion was always based on the inclusion of the oldest study and not on the largest number of effect sizes. In addition, we conducted several sensitivity analyses – including analyses conducted on sets without influential cases (Viechtbauer, 2010; Viechtbauer & Cheung, 2010) – to mitigate concerns about heterogeneity-related false positive rates of bias indication (see section 'Thematic Grouping and Sensitivity Analyses Based on Measures of Heterogeneity and Influential Cases').

We extracted the following information from each meta-analysis that was eligible for inclusion: Publication year, covered timespan, meta-analysis type (Hedges & Olkin vs. Hunter & Schmidt), reported number of effect sizes and participants included in the meta-analysis, reported meta-analytic summary effect and standard error, reported measures of heterogeneity (I^2, τ^2), reported initial study effect size and its corresponding *p*-value, reported initial study sample size, assessment of publication bias (yes/no), and the publication bias detection methods that had been used (direct comparison/failsafe N/visual inspection of funnel plot/trim-and-fill analysis/Begg & Mazumdar's rank test/Sterne & Egger's regression/selection models/PET-PEESE/Test of Excess Significance/p-curve/p-uniform/cumulative meta-analysis/other non-standard method). Moreover, in meta-analyses that used at least one bias detection method, we recorded if the results of these methods were indicative of no (defined as authors concluding that no bias was present in their datasets or giving no conclusion about bias apart from descriptively presenting results), some (authors suggested that the presence of bias did not influence their results or bias was only present in some of the analyzed distributions or methods), or considerable bias (authors concluded that publication bias was present in their meta-analysis).

If the data had been reported in the text of publications or their supplementary files, we extracted the primary data reported within meta-analyses[1]. In case of a Hunter and Schmidt-type meta-analysis, we extracted observed instead of artifact-corrected primary study and summary effect sizes. In nine cases, we coded the corrected instead of the observed meta-analytic summary effect (not used in analyses and coded for descriptive purposes only; Badura et al., 2020; Ben-Shakhar & Elaad, 2003; De Dreu & Weingart, 2003, DeChurch et al., 2013; Greer et al., 2018; Martocchio & Oleary, 1989; Premack & Wanous, 1985; Riketta, 2008; Shockley et al., 2017). The extraction of observed primary study effect sizes was not possible only in one case (De Dreu & Weingart, 2003), where we extracted corrected effect sizes instead. The mean difference between reported and recalculated meta-analytic summary effects was therefore trivial (z = -.008; Md = -.005, range: -.16 to .10). Discrepancies between the reported and recalculated summary effect exceeding a small effect ($z > |.10|$; $k = 4$) were mostly attributable to differences in weighting (sample size weighted means in original analyses versus inverse-variance weighting using random effects models in our reanalyses; Ng & Feldman, 2015; Cao & Drasgow, 2019; Robertson & Downs, 1989). Only in the case of the oldest meta-analysis in our sample that provided primary data (Steel & Ovalle, 1984), we were not able to identify the source of the discrepancy between reported and recalculated summary effect ($z_{\text{diff}} = .10$).

In addition, we manually coded the publication status of each effect size dichotomously into published (journal articles, book chapters, or books) or unpublished sources (conference presentations, doctoral dissertations, unpublished manuscripts, raw data, and technical or internal reports) by using the information provided in the reference lists of the respective meta-analyses.

Statistical Analyses

An overview of our analyses on the meta-analytic and meta-meta-analytic level is provided in Figure 1.

Analyses on the Meta-Analytic Level

Preparation of Primary Data. Prior to our meta-analytic calculations, we converted effect sizes and corresponding standard errors from all meta-analytic datasets into a common effect size metric, (i.e., Fisher *z*) using standard formulas as outlined in Borenstein et al. (2009). Seventy-two meta-analyses eligible for reanalysis (i.e., primary data available and including at least 10 independent effect sizes) reported Pearson *r* as their original metric. Another thirteen reported Cohen *d* or Hedges *g* but did not provide corresponding standard errors or confidence intervals for primary study effect sizes. Only two studies provided effect sizes as well as standard errors in a metric other than Pearson *r* (Hedges *g*; Keith & Freese, 2008, and Cohen *d*; Paustian-Underdahl et al., 2014). For consistency and because all effect sizes had to be transformed to Pearson *r* prior to conversion in Fisher *z*, all corresponding standard errors were calculated as 1 / sqrt(n – 3). Dependencies in primary datasets were resolved by computing a single, sample size-weighted effect size that was aggregated on the lowest level of independence (i.e., sample or study level).

Statistics Calculated for Meta-Analytic Datasets. We calculated several summary and publication bias statistics for each of these meta-analytic primary datasets. First, we recalculated the meta-analytic summary effect and its corresponding standard error. Specifically, we used random-effects models with inverse variances as study weights and estimated the between-study variance via maximum likelihood estimation.

Second, we applied nine standard and more modern bias detection methods that are based on different methodological rationales in each dataset. We purposefully used a large number of heterogeneous publication bias detection methods because evidence from several meta-analytic applications (Pietschnig et al., 2019; van Aert et al., 2019) as well as simulation studies (Carter et al., 2019; Renkewitz & Keiner, 2019) have shown that different bias detection methods are not equally sensitive to different publication bias scenarios and sources. As outlined below, varying degrees of heterogeneity, bias severity, or number of effect sizes are only some of the characteristics that may differentially affect the sensitivity of publication bias detection methods. While meta-analysts interested in the investigation of publication bias in their specific set of effect sizes may choose and discuss methods that are most appropriate under the very specific condition of their data (e.g., based on power considerations), this is not an option in our meta-meta-analysis. The large number of datasets that we examine in our study differ in their characteristics and potentially confounding bias may be expected to be due to various causes. Therefore, and in line with other recent meta-meta-analyses investigating publication bias (Pietschnig et al., 2019) as well as current recommendations (e.g., Coburn & Vevea, 2015; Kepes et al., 2012; van Aert et al., 2019), applying a broad array of detection methods to all sets of effect sizes was deemed to be a reasonable strategy in our analysis.

We applied nine standard and more modern bias detection methods to our datasets: Begg & Mazumdar's rank test (Begg & Mazumdar, 1994), Sterne & Egger's regression test (Sterne & Egger, 2005), PET-PEESE (Stanley & Doucouliagos, 2014), the trim-and-fill approach (Duval & Tweedie, 2000a, 2000b), the test of excess significance (Ioannidis & Trikalinos, 2007), selection models according to Vevea and Woods (2005), *p*-curve (Simonsohn et al., 2014), *p*-uniform (van Assen et al., 2015), and *p*-uniform* (van Aert & van Assen, 2018). We provide a detailed description of each method, including discussions of their strengths and weaknesses, as well as a description of outdated or inappropriate but popular methods (i.e., the failsafe-*N*, a direct comparison of published and unpublished effect sizes, and the visual inspection of the funnel plot) in the Supplement S1.

Thresholds and Conceptualization of Bias Indication. Our assumed thresholds for bias indication (see Table 1) follow common guidelines and meta-meta-analytic applications (e.g., Kepes et al., 2012; Pietschnig et al., 2019). To arrive at an omnibus indication of publication bias within a meta-analysis, we employed two approaches that have been used in the literature: First, we calculated a dichotomous indication of bias using degree-based methods of publication bias (i.e., stemming from the trim-and-fill procedure, PET-PEESE, a moderate one-tailed selection model, and p-uniform*, see Figure 1), as well as the maximum difference between a leave-one-out analysis and the meta-analytic summary effect (to account for potential outliers). This approach is particularly common in publication bias investigations in the field of I/O Psychology (e.g., Field et al., 2020; Kepes & McDaniel, 2015) and is based on the calculation of two measures of distance (i.e., the baseline range estimate, BRE, and the maximum range estimate, MRE) between the meta-analytic and the bias-adjusted summary effects. The BRE is defined as the maximum absolute difference between the meta-analytic summary effect and any bias-adjusted summary effect. The MRE is defined as the maximum absolute difference between any two of the bias-adjusted summary effects. Following recommended benchmarks and empirical applications (Field et al., 2020; Kepes et al., 2012), bias indication was considered to be non-negligible if both the BRE and the MRE exceeded 20% of the meta-analytic summary effect. A graphical display of this approach (Field et al., 2020) grouped by research topics within I/O Psychology (Cascio & Aguinis, 2008) is provided in the Supplement S2.

Second, we counted the number of methods indicating publication bias per study (Pietschnig et al., 2019). Of note, several of the methods that we applied to the meta-analytic datasets are based on the same underlying rationale, for example the assessment of small study effects (trim-and-fill analysis, Sterne & Egger's regression, Begg & Mazumdar's rank test, PET-PEESE) or the distribution of *p*-values (*p*-curve, *p*-uniform). Thus, the number of methods indicative of bias might be confounded with the number of methods that are based on a certain rationale. In the set of methods that we used, four methods are based on the rationale of assessing small study effects. For this reason, we considered a study to be indicative of severe bias in the discussion of results only in case of at least five methods being indicative of bias (i.e., ensuring at least one method not assessing small study effects indicates bias). Frequencies for bias indication by studies are presented with the possibility of correlated methods in mind. The resulting dichotomous variable (i.e., bias is negligible vs. non-negligible based on the BRE/MRE approach) and count variable (i.e., number of bias detection methods indicative of bias) were then used as dependent variables in unweighted logistic and Poisson meta-meta-regressions to assess predictors of bias indication (see Figure 1 and section 'Analyses on the meta-meta-analytic level').

Analyses on the Meta-Meta-Analytic Level

All Meta-Analyses. For all eligible studies (regardless of primary data availability), we first examined if authors had used any method to empirically detect publication bias and if so, how many and which bias detection methods had been used. In addition, we investigated changes in the use of publication bias detection over time in two ways: First, we investigated whether the number of meta-analyses that report publication bias analyses had changed over time. In this vein, we assessed associations between the use of any publication bias detection method in a meta-analysis (0 = no, 1 = yes) and its publication year by means of a point-biserial correlation. These correlations were calculated both for the use of any bias detection method as well as for each method separately.

Second, we investigated whether the number of methods within meta-analyses that are used to detect publication bias had changed over time. We did so by calculating Spearman's rank correlation coefficient between meta-analytic publication year and the number of methods used within a meta-analysis. Moreover, we visually inspected an exploratorily fitted loess curve to the bivariate scatterplot. We then formally investigated a potential change in the strength of the linear association between publication year and number of bias detection methods by fitting a segmented line regression to the data. Our model was based on a Poisson regression with one breakpoint and we tested the significance of the difference in slopes using a Score-type test (Muggeo, 2016). This approach allowed us to empirically determine if there is an identifiable time point that marks a change in bias detection methods application practice.

Meta-Analyses with Primary Data. Subsequently, we synthesized the meta-analytic results that we had obtained from reanalyzing primary datasets as described above on a meta-meta-analytic level. Although some authors use a more lenient threshold (e.g., van Aert et al., 2019), all analyses are based on primary datasets that included at least ten effect sizes, following well-established recommendations (e.g., Sterne et al., 2011).

In addition, we ran all our analyses twice, once using all meta-analyses and once using only relatively homogenous meta-analyses from the lowest quartile of the τ^2-range within our meta-meta-analytic sample. This was done because many publication bias detection methods have been shown to exhibit undesirable properties when meta-analyses show moderate-to-large heterogeneity (e.g., Ioannidis & Trikalinos, 2007; van Aert et al., 2016) or were explicitly designed under the assumption of a fixed underlying true effect (see e.g., Renkewitz & Keiner, 2019).

We intended to determine the homogeneity of the eligible meta-analyses by assessing whether the range of their respective prediction intervals for the summary effect yielded values smaller than $r = |.11|$ (i.e., following the approach of Koslowsky & Sagie, 1993, who recommended this value for the credibility interval in Hunter and Schmidt-typed meta-analyses to identify possible non-trivial moderator effects). Even though we followed the Hedges and Olkin approach in our study, this approach makes sense because prediction and credibility intervals can be considered to be virtually equivalent (Borenstein et al., 2009) and Koslowsky and Sagie (1993) only controlled for sampling error (i.e., similar to a random-effects model in the Hedges and Olkin approach) but no other statistical artifacts in their simulation study. However, only two of the meta-analyses in our re-analysis yielded values that were smaller than this cut-off (Sackett et al., 2017; Stewart & Roth, 2001), thus precluding further meta-meta-analytic analyses. Consequently, we selected a comparatively homogenous meta-meta-analytical subset by means of a relative cutoff (i.e., meta-analyses in the bottom quartile of the τ^2-range).

Moreover, we ran all our publication bias analyses within meta-analyses both on the full set of effect sizes and then on published effect sizes only (see e.g., Harrison et al., 2017; O'Boyle et al., 2014, for a similar approach). Therefore, we report a total of four meta-meta-analytic outcomes (i.e., resulting from the above described 2 x 2 combinatorial possibilities of overall in terms of variability vs. in the lowest quartile of the τ^2-range and overall in terms of publication status vs. published-only effect sizes).

For all four of these approaches, we first examined the average between-method agreement of bias detection methods by calculating pairwise phi coefficients. For descriptive purposes, we also calculated the meta-meta-analytic summary effect for all four approaches by synthesizing our recalculated meta-analytic summary effects using a random-effects model with maximum likelihood estimation and the inverse of the squared standard errors of the meta-analytic summary effects as weights.

Second, we investigated predictors of bias investigation using theory-guided unweighted hierarchical stepwise logistic as well as Poisson meta-meta-regressions. Predictors were standardized and entered in three steps for two different outcomes, namely a dichotomous indication of bias based on the combined BRE/MRE (0 = negligible amount of bias, 1 = non-negligible amount of bias; logistic regression model), as well as the number of methods that indicated bias (Poisson regression model). For both outcomes, we entered the absolute meta-analytic summary effect size in a first step, the absolute initial effect size in a second step, and the number of effect sizes as well as measures of heterogeneity (I^2 and τ^2 in two alternative models) in a third step as predictors. We ran all regression models separately for bias indications based on the full set of effect sizes as well as on the subset of published effect sizes only. We did not calculate regression models for the subset of meta-analytic sets in the lowest quartile of the τ^2-range because of low case numbers (k = 22 for the full sets and k = 21 for the published sets).

Thematic Grouping and Sensitivity Analyses Based on Measures of Heterogeneity, Artifact Correction, and Influential Cases

Given the topical heterogeneity of meta-analyses published in the *Journal of Applied Psychology* (and a possibly differential susceptibility to bias across topics, see e.g., Kepes et al., 2014), we also grouped all meta-analyses according to major topics within the field of I/O Psychology. To this end, we used the taxonomy outlined in Cascio & Aguinis (2008) that identifies 15 broad topical areas within the field. Figure 1 outlines all analyses that were conducted based on this stratification. Results of these analyses are provided in the Supplements S3 to S7.

It is well-established that heterogeneity in meta-analytic datasets can cause false positive rates in several publication bias detection methods (Vevea et al., 2019). While we provide an analysis of relatively homogenous meta-analytic datasets by running all our analyses on sets from the lowest quartile of the τ^2-distribution, we cannot rule out that heterogeneity in these studies still poses a threat to the credibility of publication bias methods used. This concern is further exacerbated by the fact that the range of the prediction interval did not exceed $r = |.11|$ in only two meta-analyses eligible for re-analysis (see Section 'Meta-Analyses with Primary Data'). Thus, we tried to mitigate this limitation in several other ways: First, we present all results for meta-analyses with an $I^2 < 25\%$ as a relative measure of homogeneity (see Supplements S8 to S10; k = 3 for full and k = 4 for published sets).

Second, we re-ran all of our reanalyses on meta-analytic datasets excluding influential cases (see Figure 1 and Supplements S11 to S16), defined as any of eight different approaches to outlier diagnostics according to Viechtbauer and Cheung (2010; see also Viechtbauer, 2010, and Field et al., 2020) being indicative of an outlier (default setting in the implementation using the *metafor* package in R, Viechtbauer, 2010). This was done because outliers or influential cases might introduce additional heterogeneity within the data (Lin et al., 2017), which in turn could unduly influence publication bias detection methods. Removed effect sizes from meta-analyses with at least ten effect sizes ranged from 0 to 5 (M = 0.74, Md = 1).

Third, we also provide sensitivity analyses using artifact-corrected instead of observed primary study effect sizes to account for additional heterogeneity introduced by measurement error. Specifically, one author [MS] extracted reliability information (r_{xx}, r_{yy}, or both) for all meta-analyses that were based on the Hunter and Schmidt-typed approach. To ensure interrater reliability as well as coding accuracy, a second experienced coder [JP] extracted reliability information from 20% (k = 9) studies (coding agreement 97.9%). The remainder of the studies was rechecked by the first coder in another turn for accuracy (coding agreement 99.8%). Discrepancies and ambiguous cases were resolved through discussion by the researchers.

In all, reliability information from 46 meta-analytic sets could be extracted, with 45 of them being suitable for re-analysis (i.e., providing at least ten effect sizes). We corrected the observed effect sizes and their corresponding sampling variances by dividing them by the square root of the product of r_{xx} and r_{yy} (effect sizes) and the product of r_{xx} and r_{yy} (sampling variances), respectively. Whenever only artifact distributions were provided, we imputed the distribution means for r_{xx} and r_{yy}. When objective (i.e., perfectly reliable) measures were used in primary studies, r_{xx} or r_{yy} were set to 1, following the approach as reported in the respective meta-analysis. The mean difference between the reported and the re-calculated summary effect sizes was z = 0.045 (range = 0 to 0.166). Minor differences were to be expected given that some studies may have employed additional artifact corrections (e.g., range variation adjustments) and – consistent with our main analysis – we used the Hedges and Olkin approach (rather than the Hunter and Schmidt approach) for assigning study weights in our syntheses. This approach enabled us to assess potential differences in outcomes of publication bias detection methods depending on the use of observed or artifact corrected datasets.

In four cases (Hong et al. 2013; Ng & Feldman, 2015; Steel & Ovalle, 1984; Casper et al., 2018), the difference exceeded $z = |.10|$. When re-calculating the summary effect based on the Hunter and Schmidt approach in *metafor*, this difference was reduced to a trivial amount (i.e., $z < |.10|$) in three cases. Only for Steel and Ovalle (1984) and similar to our main analysis, the cause of the difference could not be identified. Subsequently, we employed all nine publication bias detection methods on corrected sets and compared results between corrected and corresponding observed datasets. Results of these analyses can be found in Supplements S17 to S19.

Software and Data Availability

All statistical analyses were performed using the open-source software R. Data, R code, and all Supplements (S1 to S23) are provided at https://osf.io/dqc3y/.

Results

Out of the 362 hits that had been retrieved by our literature search, 128 meta-analyses were eligible for inclusion in our study (Figure 2). Descriptive statistics on the meta-analytic and meta-meta-analytic level for all sets of meta-analyses (all vs. only homogenous sets) and effect sizes (all vs. only published effect sizes) are reported in Table 2. In all, our meta-meta-analysis includes 128 meta-analyses (publication years: 1982 to 2020) comprising 7,263 effect sizes and 3,209,663 participants. One hundred seventeen meta-analyses (91%) employed a Hunter and Schmidt-typed approach and eleven (9%) used the Hedges and Olkin method.

Eighty-seven meta-analyses (4,988 recalculated effect sizes; 2,801,851 participants) provided primary data that made reanalyses possible. Meta-analyses in the lowest quartile of the τ^2-range ($k = 22$ for full sets; $k = 21$ for published sets) comprised 1,452 effect sizes (896 for published sets) and 1,536,793 participants (published sets: 603,428). The absolute meta-meta-analytic summary effect calculated from these datasets was small-to-moderate in size, yielding $z = 0.26$ for both full [0.22; 0.30] and published [0.23; 0.30] sets. Interestingly, the meta-meta-analytic summary effects calculated from sets in the lowest quartile of the τ^2-range were considerably smaller in size ($z = 0.14$ [0.10; 0.18] for full sets and $z = 0.16$ [0.11; 0.20] for published sets). As expected, the absolute meta-meta-analytic summary effect for corrected datasets ($k = 45$; not shown in Table 2) was somewhat larger (z = 0.31 [0.26; 0.37]) than the corresponding observed summary effect ($z = 0.26$ [0.21; 0.31]). Relative heterogeneity in primary datasets as measured by the I^2-statistic was high across all four main sets, including those in the lowest quartile of the τ^2-range (average I^2 = 77%–82% weighted, 60%–77% unweighted).

All Meta-Analyses

Bias Detection Prevalence

Table 3 shows frequencies of the publication bias detection method use over time in the *Journal of Applied Psychology*. Out of the total 128 examined meta-analyses, 53 (41%) indicated that publication bias had been investigated. Within these meta-analyses, on average 1.74 methods (*Mdn* = 1, range = 1 to 7) were used to assess bias. As we outline in our introduction, it needs to be acknowledged that some of these methods were invented or popularized in recent years only. Thus, the following overall results should be interpreted with the historical emergence and uptake of these methods in mind. The most frequently used method was a direct comparison of effects from published and unpublished sources (used in 24 meta-analyses, representing 45% of meta-analyses that used any form of bias assessment), followed by the fail-safe *N* approach ($k = 20$, 38%), visual funnel plot inspection, and the trim-and-fill-analysis (both $k = 15$, 28%). Sterne and Egger's regression approach was used in four (8%), and Begg and Mazumdar's rank test as well as a cumulative meta-analysis were used in three (6%) meta-analyses respectively. The test of excess significance, *p*-curve, and selection model approaches were used in one meta-analysis each (2%). Other standard publication bias assessment methods (i.e., PET-PEESE, *p*-uniform) were not applied in any meta-analysis in our sample.

Five meta-analyses applied non-standard forms of bias assessment. These included a correlation between effect size and sample size ($k = 2$), a fail-safe *N* variant according to a now outdated version from Hunter and Schmidt (1990; i.e., a fail-safe number that is based on participant instead of primary study number; $k = 1$), a direct comparison between effect sizes that had been published in a test manual and those that had been published elsewhere ($k = 1$), and an approach resembling Begg & Mazumdar's rank test, although not reported as such and using a different *p*-value threshold ($k = 1$).

Time Trends in Bias Assessment

We observed a significant but small change over time in the number of studies that reported the results of at least one publication bias detection method ($r_{pb} = .19$ [.02; .35]; see, rightmost column of Table 3). This change over time seems to be particularly driven by an increase in bias detection in the last decade (2011–2020; 51%, vs. .27% and 32% in previous periods). In addition, we found an increase in the number of methods that were used to detect publication bias ($r_s = .33$ [.16; .47] overall and $r_s = .53$ [.31; .70] in those meta-analyses that investigated bias). This means that in an addition to a somewhat increased use of bias detection methods, particularly during the last decade, those meta-analysts that did investigate bias did it more thoroughly.

With regard to the use of different publication bias detection methods, we found several significant time trends in both directions (see Figure 3 as well as rightmost column of Table 3). On the one hand, the use of fail-safe *N* ($r_{\mathrm{pb}} = -.54$[-.71; -.31]) decreased over time, indicating a decrease of the popularity of this method, most likely because of its inadequate evidential value. On the other hand, the trim-and-fill analyses ($r_{pb} = .46$ [.22; .65]), as well as the visual funnel plot inspections ($r_{pb} = .42$ [.17; .62]) increased in use over time.

We formally investigated the change in the strength of the time trend for the number of bias detection methods used as indicated by the loess curve (Figure 4) and by means of a segmented line Poisson regression with one breakpoint. Visual inspection of the loess regression suggests an increasing strength of the investigated association around the year 2010. This interpretation is supported by the observed significant change in slopes ($\Delta b = 0.16$, $p < 0.01$) in our segmented line regression with a breakpoint in 2010. Whilst there was no significant association in meta-analysis publication year and the number of their reported bias detection methods prior to 2010 ($b_1 < 0.01$, 95% CI [-0.05; 0.07]), a significant positive association was observed thereafter ($b_2 = 0.17$, 95% CI [0.09; 0.25]).

Bias Conclusions in Meta-Analyses

When counting the bias conclusions (no vs. some vs. clear indication of bias) from the meta-analyses that had reported results from at least one detection method ($k = 53$) we observed that 77% reported no concerns ($k = 41$) and 19% ($k = 10$) indicated some concern about bias. Only in two meta-analyses (4%), the authors reported that they had observed a clear indication of bias.

Meta-Analyses with Primary Data

Prevalence of Publication Bias

Bias Indication Differences between Methods. Table 4 shows how often individual detection methods indicated bias according to our predefined thresholds when we applied them on the primary data of the meta-analyses. The frequency of bias indication varied considerably between methods: Most indications in the full set of meta-analyses resulted from funnel-plot-asymmetry-based methods, such as PET-PEESE (45% of datasets), Sterne and Egger's regression (25%), Begg and Mazumdar's rank test (22%), and trim-and-fill (21%). In addition, the moderate one-tailed selection model approach indicated bias in 35% of sets. Bias detection methods based on *p*-values yielded considerably fewer bias indications in our reanalyses (from 0% when using *p*-curve to 7% when using *p*-uniform*).

Analyses that were based on published effect sizes only showed a similar pattern. Methods investigating small study effects indicated bias from about 17% (Begg & Mazumdar's rank test) to 51% of cases (PET-PEESE), the selection model approach in 32% of cases, whilst *p*-value-based methods yielded indications in 4% of cases (*p*-uniform and *p*-uniform*) at most. Results of excess significance tests were about half-way in-between funnel-plot asymmetry-based methods and *p*-value-based methods in both subsets.

The same rank order was also found when examining datasets in the lowest quartile of the τ^2-range (k = 22 for full and k = 21 for published sets), with funnel-plot asymmetry-based methods and the selection model approach yielding the highest indication of bias. This general trend (high share of bias indication in funnel-plot-asymmetry based methods and the selection model approach, low share in *p*-value based methods) was also found when grouping meta-analyses by topic (see Supplement S5), when analyzing datasets without influential cases (see Supplement S12), and when analyzing datasets with corrected effect sizes (see Supplement S17). Interestingly, prevalences of bias indication differed only to a trivial degree (3% at most) in observed and corrected sets of effect sizes (k = 45; see also Supplement S17). In homogenous sets as defined by the I^2-statistic (k = 3 for full sets and k = 4 for published sets) only PET-PEESE (2 out of 3 and 1 out of 4, respectively) and Sterne and Egger's regression (1 out of 3 and 1 out of 4 respectively) indicated bias in any meta-analysis (see Supplement S9).

Bias Indication Differences between Meta-Analyses. Table 5 provides frequencies of meta-analyses in which detection methods were indicative of bias. We found no bias indication in about 31% to 38% of meta-analyses in full sets and published sets, and 23% to 29% in sets from the lowest quartile of the τ^2-range. Again, the prevalence of no bias indication by study did not differ between corrected and observed sets (31% and 33%; see Supplement S18).

These numbers are broadly in line with bias indication according to the BRE/MRE-measure that indicated no bias in 23% to 31% of cases in main analyses (Table 5), in 33% to 36% of cases in sets without influential cases (Supplement S13), and in 1 out of 3 and 1 out of 4 cases in homogenous sets according to I^2 (Supplement S10). Interestingly, no bias indication according to BRE/MRE was somewhat higher in observed sets of effect sizes (36%) than in corrected sets of effect sizes (22%; Supplement S18).

Severe bias indication (as defined by at least five methods indicating bias), was found in 7% to 8% in full and published sets (Table 5; similarly 8% for sets without influential cases, Supplement S13), 0% to 14% in sets from the lowest quartile of the τ^2-range (Table 5), 0% in homogenous sets according to I^2 (Supplement S10), and did not differ between corrected and corresponding observed sets (4% to 7%; Supplement S18).

When stratified by research topic (and with the caveat of low case numbers per set), no indication of bias was found in 0% (Decision Making, Training and Development [full set]) to 57% (Research Methodology and Psychometric Issues [published set]) of meta-analyses within sets (see Supplement S6 and S7) Similarly, negligible bias indication according to the BRE/MRE measure ranged from 0% (Decision Making, Training and Development) to 63% (Career Issues). Conversely, severe bias indication was found in 0% (Career Issues, Decision Making [published set], Leader Influence [full set], Performance Measurement and Work Outcomes, Research Methodology and Psychometric Issues) to 33% (Decision Making [full set]) of meta-analyses within sets. Of note, 33% of meta-analyses also indicated severe bias in the set of meta-analyses grouped in a miscellaneous ‘Other’ category (published set). This category includes among others meta-analyses pertaining to stereotype threat (Nguyen & Ryan, 2008; Shewach et al., 2019), a phenomenon that has been shown to be prone to publication bias (Zigerell, 2017).

Agreement between Methods

Pairwise phi coefficients for agreement between the bias detection methods and for every set of meta-analyses and effect sizes can be found in Supplementary Tables S14 (sets without influential cases), S20 (lowest quartile of the τ^2-range), S21 (main full and published sets), and S19 (corrected and corresponding observed effect sizes).

Predictors of Bias Indication

Results of the final regression models for both the logistic and the Poisson regression are reported in Tables 6 and 7 (details of the full stepwise regression models are reported in the Supplementary Tables S22 and S23, as well as in S15 and S16 for sets without influential cases). In our logistic regression models (Table 6), the strength of the meta-analytic summary effect was significantly and negatively associated with bias indication according to the combined BRE/MRE-measure regardless of effect size sets or measure of heterogeneity (range ßs = -0.94 to -1.65). Initial effect strength was significantly and positively associated with bias indication in the published set using I^2 as a measure of between-study heterogeneity (ß = 0.98, p = .014). In sets without influential cases (Supplement S15), a similar picture emerged: The meta-analytic summary effect was negatively related to bias indication in all four final models (range ßs = -0.93 to -1.70). Interestingly, initial effect strength was significantly positively associated with bias indication in all sets (range ßs = 0.70 to 1.07).

In our Poisson regression models (Table 7), both the meta-analytic summary effect and the initial effect predicted the number of methods that were indicative of publication bias. The meta-analytic summary effect was again negatively associated with the number of methods indicating bias (range ßs = -0.65 to -0.92), the strength of the initial effect was positively related to the number of bias indicative methods (range ßs = 0.23 to 0.37). In sets without influential cases (Supplement S16), the negative association between the meta-analytic summary effect and number of methods indicating bias was also found in all four final models (range ßs = -0.61 to -1.01). The positive association between initial effect strength and number of methods indicating bias was found in final models for published sets only (ß = 0.30 and 0.25, respectively).

The number of effect sizes within the respective meta-analyses predicted number of bias indicative methods only in the Poisson regression model (full set) and when using τ^2 as a measure of between-study heterogeneity (ß = 0.14, p = .039). Across all sets (including those without influential cases) and models, τ^2 was significantly and positively related to bias indication.

Discussion

In this meta-meta-analysis, we showed that indications of publication bias are a common occurrence in I/O Psychology that often remain undetected due to the underuse of appropriate detection methods. This seems to be rooted in two related phenomena: First, in only about two fifths (41%) of meta-analyses in the *Journal of Applied Psychology*, and therefore conceivably in I/O Psychology in general, have any detection methods been applied at all. While we showed that both the overall use of bias detection methods as well as the number of used bias detection methods have increased over time, still only about half (51%) of meta-analyses have engaged in bias detection during the last decade.

Second, in meta-analyses that used any form of bias detection method, it is more often than not (i) a single and (ii) an outdated or (by current standards) inappropriate method to detect publication bias. This has serious ramifications for the evidential value of the field of I/O Psychology as a whole, because effect sizes on the primary study and consequently the meta-analytic level may be taken at face value when many of them are, in fact, inflated.

Publication Bias Detection Prevalence in I/O Psychology Only Increases Slowly

The with 41% low overall prevalence of publication bias detection found in our study (contrasting about 70% in Psychology in general; Banks et al., 2012; Ferguson & Brannick, 2012) corroborates concerns that have been voiced in I/O Psychology (Banks et al., 2012; Harrison et al., 2017; Kepes et al., 2012) about the comparatively little attention in this field that has been dedicated to potential influences of publication bias and the possibilities for its detection. Indeed, for the period between 2001 and 2010, our findings of 27% of meta-analyses using at least one bias detection method are consistent with those of Kepes et al. (2012; see also Banks et al., 2012), who reported an overall publication bias detection prevalence of 31% in four top-tier I/O journals from 2005 to 2010 as well as no changes in the number of meta-analyses that report results from bias detection methods over this time.

Our findings are only partially in line with those of Aytug et al. (2012) who investigated meta-analytic reporting practices in 11 top-tier I/O journals (1995–2008) and found a slightly lower publication bias detection prevalence (17.7%), but a significant and positive time trend (r = .18)[2]. The present study extends the findings of both of the available accounts by providing evidence about the prevalence of publication bias detection in I/O Psychology in more recent (2011–2020) as well as prior (1982–1994) time periods.

This considerably larger timeframe gives rise to cautious optimism: We found that the use of any publication bias detection method has increased over time, albeit to a small degree, r_{pb} = .19. During the past decade (2011-2020), publication bias detection has been used in about half of meta-analyses (51%) and in most recent years (2019-2020) this number has risen to about two thirds (67%). This suggests that the calls for higher awareness regarding publication bias (Kepes et al., 2012) – coupled with improved meta-analytic reporting guidelines (e.g., the Meta-Analysis Reporting Standards, MARS; American Psychological Association, 2010) – may have indeed had an impact on current reporting practices and awareness in authors, reviewers, and editors about publication bias in I/O-related meta-analyses. However, even when leaving concerns about the appropriateness and number of applied methods aside (see below), it should be noted that a third of published meta-analyses in this field did not report a single method of bias detection in most recent years, when these guidelines were already well in place.

Differences in the application and prevalence of publication bias detection methods in different subfields (here: I/O Psychology vs. general Psychology) are not altogether surprising because they are characterized by different research traditions. For example, more than 90% of the presently investigated meta-analyses used a Hunter and Schmidt approach, a method that focuses on alleviating potentially effect-deflating effects due to statistical artifacts such as unreliability or range variation (Hunter & Schmidt, 2004). In many other Psychology subfields outside of I/O Psychology, the Hedges and Olkin (1985) approach – typically focusing more on potential effect-inflation – seems to be more popular (e.g., used in 76% to 98% of meta-analyses in Personality and Social Psychology as well as multidisciplinary Psychology; Siegel et al., 2019). In addition, most available software packages to date that are used to conduct Hunter and Schmidt-typed meta-analyses do not include routines to address publication bias (excepting cumulative meta-analysis, Schmidt & Hunter, 2015). Whereas current publication bias detection methods have been developed under the Hedges and Olkin paradigm, publication bias detection methods based on psychometric meta-analysis have yet to be developed, arguably owing to the complexity of this method. In all, it seems that publication bias detection in Psychology may be subfield-dependent and contingent on different meta-analytic schools, journal traditions, authoritative sources, and software applications in the conduction of meta-analyses (Kepes et al., 2012).

The Number of Used Bias Detection Methods Increases (but not all Methods are Appropriate)

We observed clear increases in the number of used methods over time. Interestingly, this association only emerged by the year 2010, almost coinciding with the advent of the confidence crisis in 2011. This means that in addition to (slightly) more authors being concerned about publication bias, those who are, investigate it more thoroughly. On a first glance, these results are encouraging. However, a closer look reveals that there remains considerable room for improvement, which can be illustrated by two observations.

First, even though, on average, more bias detection methods are used, the median number of methods remains two. This means that more than half of meta-analysts (with the intent to detect publication bias) still rely on only a single or at most two methods. As we have shown here and was also shown elsewhere (Carter et al., 2019; van Aert et al., 2019), this is problematic because publication bias detection methods vary in their sensitivity and specificity in bias detection under different scenarios and conditions, and consequently in their agreement with one another. Therefore, and because no method to date can safely be recommended over all others in every scenario, current recommendations advise researchers to aim for triangulation of publication bias detection methods (e.g., Vevea et al., 2019).

Second, we observed clear evidence for changes in the use of certain detection methods. Whilst the application of fail-safe *N*s decreased over time, visual funnel plot inspections and applications of the trim-and-fill method increased. In addition, a direct comparison between published and unpublished effect sizes remained the most popular method throughout the last two decades. Again, it seems encouraging that the use of an outdated method such as fail-safe *N* seems to have lost popularity among meta-analysts. However, our evidence suggests that this method has been predominantly replaced by other problematic methods (with the arguable exception of trim-and-fill). Indeed, two out of the three most commonly used methods in the past decade, namely the direct comparison as well as the visual inspection of funnel plots, are considered to be inappropriate to detect publication bias by current standards (Kepes et al., 2012).

In this vein, the use of several but inappropriate methods could exacerbate the problem of undetected publication bias further: Authors and readers alike might base their conclusion on the convergence of several inappropriate bias detection methods and falsely conclude that no bias is present, when, in fact, there is. Therefore, we emphasize current recommendations to rely on several bias detection methods using different methodological rationales.

Additionally, we advise applied meta-analysts to select their set of bias detection methods with appropriate regard for their meta-analytic dataset at hand.

Potential Publication Bias in I/O Psychology Often Remains Undetected

Considering the bias detection practices outlined above, it does not come as a surprise that the rate of author-reported bias indication was generally low. In 77% of cases, authors concluded that no bias was present, whereas 19% concluded that some bias was present, and only two studies (4%) found a clear indication of bias. The overall author-reported bias indication rate of 23% (i.e., representing those that reported evidence for at least some bias) that was observed in our study contrasts considerably higher author-reported rates that have been found in general psychology (41%, Ferguson & Brannick, 2012).

Arguably, author-reported rates of bias detection are influenced both by the actual amount of bias in a field (e.g., some fields or disciplines seem to be more prone to publication bias than others, Fanelli et al., 2017), as well as authors' awareness and ability to detect publication bias. The former is rooted in specific characteristics of a field (e.g., publication pressure and frequency; typical sample sizes, study designs, and consequently power considerations) and needs to be targeted at the primary study level rather than the research synthesis level. It is conceivable that I/O Psychology might indeed be less prone to publication bias than other fields (e.g., social psychology), because synthesizing effect sizes that are not focal to the primary study or from primary studies that test multiple hypotheses might be more common.

The latter, however, requires translational efforts into applied meta-analytic practice: For one, I/O Psychology has been repeatedly called out as a field that lacks awareness about publication bias and is thus particularly prone to underreporting (Banks et al., 2012; Banks & McDaniel, 2011; Harrison et al., 2017). Moreover, standard meta-analytic guidelines pertinent to the field of Psychology such as MARS (American Psychological Association, 2020) or PRISMA (Moher et al., 2009) recommend application of publication bias detection methods, but specify neither their number, nor the interpretative depth of their results. Consequently, meta-analysts may opt to descriptively present results of one (typically easy to implement) bias detection method – thus meeting guidelines' and subsequently journals' requirements – while adding little to the interpretation of their results or the detection of actual bias in their analyses.

This problem is exacerbated, as some methods – particularly those that offer an effect estimate that is adjusted for publication bias – simply lack thresholds for bias indication and depend on researchers' interpretation of the strength of effect misestimation (e.g., PET-PEESE, selection models). Moreover, a number of these methods also lack implementation in standard meta-analytic software, particularly for Hunter and Schmidt-typed analyses, because they were developed under the Hedges and Olkin framework (Kepes & McDaniel, 2015). Conversely, the methods that have been found to be popular among meta-analysts are typically easy to implement and intuitively logical (but unfortunately often inappropriate). Moreover, they offer lenient rules-of-thumb (e.g., fail-safe *N*), direct interpretability (e.g., direct comparison), or are subjective by default (e.g., the funnel plot).

Thus, applied meta-analysts need not only be equipped with methodological expertise in publication bias analyses, but also in statistical programming to perform publication bias analyses that are currently recommended. For meta-analysts applying psychometric meta-analysis, this problem is exacerbated further, because current bias detection methods are based on the Hedges and Olkin approach. Apart from theoretical considerations about the causes, correlates, and consequences of publication bias in I/O Psychology, this translates into an increased need for hands-on tutorials on publication bias methodology and their implementation, as well as increased efforts into the development of easy and free to use software solutions with clear interpretative guidelines. Web applications such as those implementing *p*-curve (http://www.p-curve.com/), *p*-uniform (https://rvanaert.shinyapps.io/p-uniform/), or selection models (https://vevealab.shinyapps.io/WeightFunctionModel/), are only some examples that help to facilitate application of more complex methods. To further enhance reproducibility and a streamlined analysis, future researchers may wish to integrate many publication bias detection methods into one single application. A recent example of such an approach is the open-source application Meta-Sen (https://metasen.shinyapps.io/gen1/), which includes several forms of outlier detection, publication bias analyses, and graphical approaches to bias triangulation (Field et al., 2020).

We further recommend preregistration of all planned publication bias detection methods in pertinent registries such as the Open Science Framework (www.osf.io) or PROSPERO (https://www.crd.york.ac.uk/prospero/) to prevent selective reporting of (favorable) publication bias analysis results. In addition, we recommend making meta-analytic datasets publicly available upon publication in order to facilitate reanalyses with different or novel methods in general and publication bias analyses in particular. This is particularly important as we have to reemphasize that meta-analysts cannot (and should not) be blamed for having used methods that are now considered to be outdated and not using those that were developed after the publication of their meta-analysis.

Indications of Potential Bias Vary Between Methods and Research Topics (but are Non-Negligible)

Whenever meta-analytic data was available, we applied nine standard and more modern methods of publication bias detection to these datasets. Importantly and considering that no publication bias detection method can currently be recommended over all others (Renkewitz & Keiner, 2019), these results should be viewed with caution, because indications of potential bias (according to the respective method) are likely confounded with shortcomings of the respective method. Thus, it is not surprising that potential bias indication varied across methods. Across sets, funnel-plot-asymmetry based methods (i.e., Sterne & Egger's regression, Begg & Mazumdar's rank test, PET-PEESE, Trim-and-fill) as well as the selection model approach consistently yielded the highest indications of bias (21% to 54%), *p*-value based methods the lowest indications (0% to 7%), and the test of excess significance fell in between. This is in line with previous studies based on simulations and real-world meta-analytic datasets that found comparatively high bias detection rates particularly for funnel-plot-asymmetry based methods (Carter et al., 2019; Lin et al., 2018; van Aert et al., 2019).

It has been repeatedly noted that false positive rates of regression-based methods, such as Sterne and Egger's regression as well as the conceptually similar PET-PEESE, tend to increase in presence of between-study heterogeneity (Rothstein & Bushman, 2012; Stanley, 2017). In our meta-meta-analysis, the mean between-study heterogeneity was indeed large (I^2 = 77% to 82%), which might have contributed to the higher detection rates found with these methods. However, the results for meta-analyses in the lowest quartile of the τ^2 -range, in sets with $I^2 < 25\%$, and without influential cases were convergent with results from the full sets, thus corroborating our findings of higher detection rates of regression-based methods, regardless of between-study heterogeneity. In addition, we found that heterogeneity due to measurement error did not influence those conclusions when comparing corrected and observed sets of effect sizes.

While exact estimates vary, the overall amount of potential publication bias detected in these meta-analytic datasets is far from trivial. For example, at least one method indicated potential bias in 62% to 77% of datasets or non-negligible bias according to the combined BRE/MRE-measure. Moreover, we found severe potential bias indication (here defined conservatively as at least five methods indicative of bias) in about 8% of datasets. These findings stand in stark contrast to the low bias rates that have been reported by the authors of the investigated meta-analyses (23%). This is indicative of an under-detection of publication bias in I/O Psychology and echoes previous concerns about the bias detection practices within this field (Kepes et al., 2012).

Although not a focus of our study, we found some evidence that different subfields within I/O-Psychology might be more prone to publication bias than others (e.g., severe potential bias indication ranged from 0% to 33% across subfields). We encourage future research to investigate differential bias susceptibility across research areas within I/O Psychology more thoroughly and in a theory-guided manner (i.e., by establishing hypotheses about why a certain subfield might be more susceptible to bias than others). This may contribute to further increase confidence in research areas within I/O Psychology that possess an established and robust evidence base and would further elucidate drivers of publication and other dissemination biases.

Publication Bias-Related Meta-Analytic and Study Characteristics Predict Potential Bias Indication

Our hierarchical logistic and Poisson meta-meta-regressions showed that the strength of meta-analytic summary effects is negatively associated with potential bias indication. This finding remained robust even when controlling for influences of between-study heterogeneity and number of effect sizes as well as in different subsets of our data.

The negative association between a meta-analytic summary effect and potential bias indication is in line with theoretical predictions. In the absence of questionable research practices such as *p*-hacking (Head et al., 2015) or HARKing (Kerr, 1998), the meta-analytic summary effect can be considered to be a good estimate of the underlying true effect size. If the underlying true effect size is large, even studies with small sample sizes should yield significant results and can therefore be considered to be publishable. Thus, selective reporting and publishing of effect sizes based on statistical significance – and therefore publication bias – is unlikely to happen in this case. In contrast, smaller true effect sizes are bound to yield smaller effect sizes in individual studies, which are typically non-significant in small samples. However, some statistical tests will reach significance due to chance, and are thereby potentially publishable. The effect sizes obtained from these analyses have to be exceptionally large to reach a nominal level of significance and represent thus a considerable overestimation of the true effect.

It should be noted that the drawback of using a relative reduction in effect strength as indication of bias (as is done in the BRE/MRE approach; the dependent variable in our logistic meta-meta-regressions) might also contribute to the negative relation between the meta-analytic summary effect and bias indication. Specifically, trivial absolute reductions in meta-analyses with a small meta-analytic summary effect are indicative of bias in the BRE/MRE approach, whereas non-trivial reductions are not considered to be indicative of bias in meta-analyses with a large meta-analytic summary effect. However, since we also found a significant association between the meta-analytic summary effect and the number of methods indicative of bias in our Poisson meta-meta-regression models, this is unlikely to be the only driver of bias indication.

In a related vein, the strength of the effect size from the oldest study included in our analysis (i.e., the initial effect) was positively associated with bias indication in our Poisson meta-meta-regression models as well as in logistic meta-meta-regression models on sets of published effect sizes. This association also remained significant when controlling for the strength of the meta-analytic summary effect as well as between-studies heterogeneity (I^2 only in the logistic meta-meta-regression) and the number of included effect sizes on the meta-analytic level. This finding is in line with the theoretical assumptions underlying the decline effect. Apart from genuine effect declines over time (e.g., a decrease in prejudicial attitudes towards an out-group), declines in effect strength may be rooted in publication-bias related mechanisms (Protzko & Schooler, 2017). Specifically, theory stipulates that initial effects that are published about a certain phenomenon might be inflated due to the selective publication of significant results, that yield – due to the often inadequate power of exploratory studies (Pietschnig et al., 2019) – also large effects. An inflated initial effect may therefore not only falsely anchor subsequent replication attempts, but it might also pollute the field considerably, as replication attempts or contradictory evidence might be less frequently cited, less visibly published, or not published at all (Greenberg, 2009; Ioannidis, 2005a; Voracek, 2014). Our observation that large initial study effects are related to bias indication within a given meta-analysis are consistent with these ideas.

While previous meta-meta-analytic investigations have empirically investigated declining effects in the field of Psychology (e.g., Nuijten et al., 2020; Pietschnig et al., 2019), we present here – to the best of our knowledge – the first empirical investigation linking the decline effect directly to publication bias indication. Well aware that this might therefore represent the starting point for a decline effect in itself (and given the non-significant results in some of our logistic regression models), we encourage researchers to replicate our findings in I/O Psychology specifically (e.g., by using a different set of journals or meta-analyses pertaining to a specific research question or tradition), but also in Psychological Science in general.

Limitations

First, we relied on meta-analyses published in a single journal, namely the *Journal of Applied Psychology*. Therefore, it is unknown to what extent our results generalize to other top-tier I/O journals or across the field of I/O Psychology. However, the *Journal of Applied Psychology* is a flagship journal and known to publish a cross-section of high-quality work pertinent to I/O Psychology. We chose this journal for two reasons: First, previous estimates of publication bias detection in this journal have been consistent with the overall estimates of bias detection rates in this field (i.e., 31%; Banks et al., 2012; Kepes et al., 2012). Second, the *Journal of Applied Psychology* does not only publish the highest share of meta-analyses among top-tier journals in I/O Psychology (Aytug et al., 2012; Schalken & Rietbergen, 2017), but it has also been found to be more transparent in several aspects of meta-analytic reporting standards and to publish meta-analyses that are cited at a significantly higher rate (Aytug et al., 2012). Therefore, we were able to gather a large sample that consists of meta-analyses that conform to often-recommended meta-analytic standards (as evidenced by the high transparency scores) and represent valuable resources for researchers in the field (as evidenced by the high citation rates). However, we encourage future research to replicate our findings based on different journals publishing meta-analyses in the field of I/O Psychology.

Second, as we have noted throughout our study, heterogeneity poses a problem to many bias detection methods (Carter et al., 2019; van Aert et al., 2019). In our meta-meta-analysis, the average between-study heterogeneity was large (I^2 = 77% to 82%) and only two meta-analyses fell below a cutoff for the presence of nontrivial moderator effects (Koslowsky & Sagie, 1993), which may have led to false positive and inflated estimates of bias indication. We alleviated this problem by rerunning all our analyses on sets that fell into the lowest quartile of the τ^2-range (k = 22), sets with I^2-values below 25% (k = 4), sets without influential cases (k = 87), and sets that were corrected for measurement error (k = 45). Despite the comparatively small number of meta-analyses in some of these sets, the key patterns of results were largely consistent with those that we found in full sets. For example, we found a virtually identical rank order of bias indication by method as well as rates of bias indication across comparatively homogenous and heterogenous sets. Additionally, results remained virtually unchanged for artifact corrected and observed sets of effect sizes, indicating that additional heterogeneity introduced by measurement error did not substantially influence the sensitivity of publication bias detection methods. However, because we cannot completely rule out that unobserved heterogeneity affected results of the publication bias detection methods that were used in our study, our results should be taken with a grain of salt. Future researchers may wish to deliberately sample homogenous meta-analytic distributions (e.g., by using the cut-off proposed by Koslowsky & Sagie, 1993) to further disentangle false positives due to heterogeneity from genuine publication bias indication.

While heterogeneity in our sample was indeed high, it is also well in line with heterogeneity estimates found elsewhere (e.g., I^2 = 74% in meta-analyses published in the

Psychological Bulletin; Stanley et al., 2018) and indicates that heterogeneity in applied settings is not only rampant, but poses a risk for biased effect estimations. Thus, we were able to apply publication bias detection methods and investigate publication bias related phenomena under real world conditions, thereby providing immediate implications with a high ecological validity for the field.

Third, when presenting our results narratively, we used a cutoff of five methods indicative of potential bias to designate instances of “severe bias”. We chose this cutoff to accommodate correlated estimates provided by the four conceptually similar methods to assess small study effects, namely the trim-and-fill procedure, Begg and Mazumdar’s rank test, Sterne and Egger’s regression, and PET-PEESE. Even if all four methods indicated bias, at least one other method needed to do so to be counted as severely biased. Arguably, this cutoff still incorporates the preponderance of publication bias methods designed to detect small study effects, both in our methodological arsenal as well as in the field of publication bias detection in general. However, the empirical correlations between those methods found in our sample (Supplement S21) were only small-to-moderate (full set; range .17 to .40) and moderate-to-large (published set; range .31 to .62) respectively. Thus, bias indication in one method did not imply bias indication in all other three methods in our sample. On the contrary, we found that correlations between other methods from different rationales exceeded those found for the four methods in question (e.g., .55 to .57 between TES and the selection model approach). In addition, these correlations do not seem to be driven by heterogeneity-related false positives, as they were of similar strength in homogenous sets (range .22 to 41 in full sets and .29 to .53 in published sets; Supplement S20).

Still, we encourage readers to employ different cutoffs (e.g., six methods) or tally the numbers of methods indicative of bias according to their own selection (e.g., choosing only one method to detect small study effects) if deemed useful. In addition, future research should aim at identifying further methods of bias triangulation that are less affected by conceptual similarity between individual bias detection methods.

Finally, we cannot rule out that there may have been differences in bias prevalence between meta-analyses whose data were published compared to those whose data were unavailable. However, no significant differences were observed in reported meta-analytic effect strength, the likelihood of using at least one bias detection method, and the number of used bias detection methods between these two subgroups. These findings render substantial confounding effects of meta-analysis characteristics due to primary data reporting practices unlikely.

Conclusion

The present meta-meta-analysis constitutes one of the most comprehensive assessments of trends and drivers regarding potential publication bias and its detection practices in the field of I/O psychology. By applying nine methods from a number of different methodological families to detect publication bias, as well as subsequently linking indication of potential bias to predictors on the primary study and the meta-analytic level in meta-meta-regressions, we provided a multi-faceted analysis of bias indication. Our study revealed under-detection of potential publication bias in the field of I/O Psychology that is rooted in (i) low publication bias detection prevalence as well as (ii) frequent applications of publication bias detection methods considered to be inadequate by current standards. However, we also observed increases in the overall use of any bias detection method as well as in the number of methods used, which gives rise to cautious optimism regarding the awareness about publication bias in the field.

On the primary study level, we emphasize longstanding recommendations to tackle the pervasive problem of publication bias, e.g., study preregistration and registries particular to I/O Psychology (Kepes et al., 2013), pre-publication replication approaches (either by self- or independent replication), and journal policies registered reports.

On the meta-analytic level, we argue for (i) increased awareness in the field of I/O Psychology with regard to publication bias, (ii) increased efforts to develop publication bias detection methods for psychometric meta-analyses, (iii) increased efforts to bridge the gap between meta-analytic methodologists and applied meta-analysts in the implementation and interpretation of modern publication bias detection methods, (iv) the application of several and state of the art bias detection methods in individual meta-analyses, (v) preregistration to prevent selective reporting of favorable publication bias analyses, and (vi) data sharing to enable publication bias reanalyses using novel or different methods.

References

References marked with an asterisk indicate studies included in the meta-meta-analysis. References marked with two asterisks (k = 87) indicate studies that also provided sufficient data for reanalysis (i.e., ten or more effect sizes). References marked with three asterisks (k = 45) were also included in analyses using corrected effect sizes.

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, *37*(1), 5–38. https://doi.org/10.1177/0149206310377113

*Allen, T. D., Eby, L. T., Poteet, M. L., Lentz, E., & Lima, L. (2004). Career benefits associated with mentoring for protégés: A meta-analysis. *Journal of Applied Psychology*, *89*(1), 127–136. https://doi.org/10.1037/0021-9010.89.1.127

***Allen, T. D., French, K. A., Dumani, S. & Shockley, K. M. (2020). A cross-national meta-analytic examination of predictors and outcomes associated with work–family conflict. *Journal of Applied Psychology*, *105*(6), 539–576. https://doi.org/10.1037/apl0000442

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). American Psychological Association.

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association. https://doi.org/10.1037/0000165-000

Augusteijn, H. E. M., van Aert, R. C. M., & van Assen, M. A. L. M. (2019). The effect of publication bias on the Q test and assessment of heterogeneity. *Psychological Methods*, *24*(1), 116–134. https://doi.org/10.1037/met0000197

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*(1), 103–133. https://doi.org/10.1177/1094428111403495

**Bachrach, D. G., Lewis, K., Kim, Y., Patel, P. C., Campion, M. C. & Thatcher, S. M. B. (2019). Transactive memory systems in context: A meta-analytic examination of contextual factors in transactive memory systems development and team performance. *Journal of Applied Psychology*, *104*(3), 464–493. https://doi.org/10.1037/apl0000329

***Badura, K. L., Grijalva, E., Galvin, B. M., Owens, B. P. & Joseph, D. L. (2020). Motivation to lead: A meta-analysis and distal-proximal model of motivation and leadership. *Journal of Applied Psychology*, *105*(4), 331–354. https://doi.org/10.1037/apl0000439

*Baltes, B. B., Briggs, T. E., Huff, J. W., Wright, J. A., & Neuman, G. A. (1999). Flexible and compressed workweek schedules: A meta-analysis of their effects on work-related criteria. *Journal of Applied Psychology*, *84*(4), 496–513. https://doi.org/10.1037/0021-9010.84.4.496

Banks, G. C., Kepes, S., & McDaniel, M. A. (2012). Publication bias: A call for improved meta-analytic practice in the organizational sciences. *International Journal of Selection and Assessment*, *20*(2), 182–196. https://doi.org/10.1111/j.1468-2389.2012.00591.x

Banks, G. C., & McDaniel, M. A. (2011). The kryptonite of evidence-based I-O Psychology. *Industrial and Organizational Psychology*, *4*(1), 40–44. https://doi.org/10.1111/j.1754-9434.2010.01292.x

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Editorial: Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology*, *31*(3), 323–338. https://doi.org/10.1007/s10869-016-9456-7

*Barsky, A., & Kaplan, S. A. (2007). If you feel bad, it's unfair: A quantitative synthesis of affect and organizational justice perceptions. *Journal of Applied Psychology*, *92*(1), 286–295. https://doi.org/10.1037/0021-9010.92.1.286

**Bauer, T. N., Bodner, T., Erdogan, B., Truxillo, D. M., & Tucker, J. S. (2007). Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology*, *92*(3), 707–721. https://doi.org/10.1037/0021-9010.92.3.707

**Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, *88*(6), 989–1004. https://doi.org/10.1037/0021-9010.88.6.989

Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning and Education*, *9*(4), 715–725. https://doi.org/10.5465/AMLE.2010.56659889

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

**Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: A meta-analytic review. *Journal of Applied Psychology*, *88*(1), 131–151. https://doi.org/10.1037/0021-9010.88.1.131

***Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, *97*(3), 613–636. https://doi.org/10.1037/a0026739

**Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology*, *96*(5), 881–906. https://doi.org/10.1037/a0023222

*Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*(2), 410–424. https://doi.org/10.1037/0021-9010.92.2.410

*Beus, J. M., Dhanani, L. Y., & McCord, M. A. (2015). A meta-analysis of personality and workplace safety: Addressing unanswered questions. *Journal of Applied Psychology*, *100*(2), 481–498. https://doi.org/10.1037/a0037916

*Beus, J. M., Payne, S. C., Bergman, M. E., & Arthur, W. (2010). Safety climate and injuries: An examination of theoretical and empirical relationships. *Journal of Applied Psychology*, *95*(4), 713–727. https://doi.org/10.1037/a0019164

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

**Brady, G. M., Truxillo, D. M., Cadiz, D. M., Rineer, J. R., Caughlin, D. E. & Bodner, T. (2020). Opening the black box: Examining the nomological network of work ability and its role in organizational research. *Journal of Applied Psychology*, *105*(6), 637–670. https://doi.org/10.1037/apl0000454

***Breuer, C., Hüffmeier, J., & Hertel, G. (2016). Does trust matter more in virtual teams? A meta-analysis of trust and team effectiveness considering virtuality and documentation as moderators. *Journal of Applied Psychology*, *101*(8), 1151–1177. https://doi.org/10.1037/apl0000113

**Burke, M. J., Salvador, R. O., Smith-Crowe, K., Chan-Serafin, S., Smith, A., & Sonesh, S. (2011). The dread factor: How hazards and safety training influence learning and performance. *Journal of Applied Psychology*, *96*(1), 46–70. https://doi.org/10.1037/a0021838

***Butts, M. M., Casper, W. J., & Yang, T. S. (2013). How important are work-family support policies? A meta-analytic investigation of their effects on employee outcomes. *Journal of Applied Psychology*, *98*(1), 1–25. https://doi.org/10.1037/a0030389

**Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: A meta-analysis examining competing theoretical models. *Journal of Applied Psychology*, *95*(1), 201–212. https://doi.org/10.1037/a0017868

Campbell, J. P. (1990). The role of theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 39–73). Consulting Psychologists Press.

**Cao, M. & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, *104*(11), 1347–1368. https://doi.org/10.1037/apl0000414

*Carr, J. Z., Schmidt, A. M., Ford, J. K., & DeShon, R. P. (2003). Climate perceptions matter: A meta-analytic path analysis relating molar climate, cognitive and affective states, and individual level work outcomes. *Journal of Applied Psychology*, *88*(4), 605–619. https://doi.org/10.1037/0021-9010.88.4.605

**Carsten, J. M., & Spector, P. E. (1987). Unemployment, job satisfaction, and employee turnover: A meta-analytic test of the Muchinsky model. *Journal of Applied Psychology*, *72*(3), 374–381. https://doi.org/10.1037//0021-9010.72.3.374

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Cascio, W. F., & Aguinis, H. (2008). Research in Industrial and Organizational Psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, *93*(5), 1062–1081. https://doi.org/10.1037/0021-9010.93.5.1062

***Casper, W. J., Vaziri, H., Wayne, J. H., DeHauw, S. & Greenhaus, J. (2018). The jingle-jangle of work–nonwork balance: A comprehensive and meta-analytic review of its meaning and measurement. *Journal of Applied Psychology*, *103*(2), 182–214. https://doi.org/10.1037/apl0000259

***Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, *83*(4), 615–633. https://doi.org/10.1037/0021-9010.83.4.615

Cheung, S. F., & Chan, D. K.-S. (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology*, *89*(5), 780–791. https://doi.org/10.1037/0021-9010.89.5.780

***Chiaburu, D. S., Oh, I.-S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, *96*(6), 1140–1166. https://doi.org/10.1037/a0024004

***Choi, C., Oh, I.-S., & Colbert, A. E. (2015). Understanding organizational commitment: A meta-analytic examination of the roles of the five-factor model of personality and culture. *Journal of Applied Psychology*, *100*(5), 1542–1567.

**Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, *94*(5), 1103–1127. https://doi.org/10.1037/a0016172

Coburn, K. M., & Vevea, J. L. (2015). Publication bias as a function of study characteristics. *Psychological Methods*, *20*(3), 310–330. https://doi.org/10.1037/met0000046

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, *80*(5), 565–579. https://doi.org/10.1037/0021-9010.80.5.565

Cooper, H., Charlton, K., & DeNeve, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*, *2*(4), 447–452. https://doi.org/10.1037/1082-989X.2.4.447

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.

**Courtright, S. H., McCormick, B. W., Postlethwaite, B. E., Reeves, C. J., & Mount, M. K. (2013). A meta-analysis of sex differences in physical ability: Revised estimates and strategies for reducing differences in selection contexts. *Journal of Applied Psychology*, *98*(4), 623–641. https://doi.org/10.1037/a0033144

*Courtright, S. H., Thurgood, G. R., Stewart, G. L., & Pierotti, A. J. (2015). Structural interdependence in teams: An integrative framework and meta-analysis. *Journal of Applied Psychology*, *100*(6), 1825–1846. https://doi.org/10.1037/apl0000027

***Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. J. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology*, *96*(3), 443–456. https://doi.org/10.1037/a0022147

*Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*, *90*(6), 1241–1255. https://doi.org/10.1037/0021-9010.90.6.1241

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F., & Pierce, C. A. (2011). Revisiting the file-drawer problem in meta-analysis. *Academy of Management Proceedings*, *2011*(1), 1–6. https://doi.org/10.5465/ambpp.2011.65869140

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology, 65*, 221–249. https://doi.org/10.1111/j.1744-6570.2012.01243.x

***De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, *88*(4), 741–749. https://doi.org/10.1037/0021-9010.88.4.741

***De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*, *101*(8), 1134–1150. https://doi.org/10.1037/apl0000110

***De Wit, F. R. C., Greer, L. L., & Jehn, K. A. (2012). The paradox of intragroup conflict: A meta-analysis. *Journal of Applied Psychology*, *97*(2), 360–390. https://doi.org/10.1037/a0024844

*Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology*, *93*(3), 685–691. https://doi.org/10.1037/0021-9010.93.3.685

***DeChurch, L. A., Mesmer-Magnus, J. R., & Doty, D. (2013). Moving beyond relationship and task conflict: Toward a process-state perspective. *Journal of Applied Psychology*, *98*(4), 559–578. https://doi.org/10.1037/a0032896

*Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology*, *87*(4), 611–628. https://doi.org/10.1037/0021-9010.87.4.611

Doucouliagos, C., & Stanley, T. D. (2013). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys, 27*, 316–339. https://doi.org/10.1111/j.1467-6419.2011.00706.x

**Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology*, *79*(4), 481–492. https://doi.org/10.1037/0021-9010.79.4.481

**Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*(5), 615–622. https://doi.org/10.1037/0021-9010.77.5.615

Duval, S., & Tweedie, R. (2000a). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

*Eatough, E. M., Chang, C.-H., Miloslavic, S. A., & Johnson, R. E. (2011). Relationships of role stressors with organizational citizenship behavior: A meta-analysis. *Journal of Applied Psychology*, *96*(3), 619–632. https://doi.org/10.1037/a0021887

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.316.7129.469

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Fanelli, D., Costas, R., & Ioannidis, J. P. A. (2017). Meta-assessment of bias in science. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(14), 3714–3719. https://doi.org/10.1073/pnas.1618569114

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, *17*(1), 120–128. https://doi.org/10.1037/a0024445

Field, J. G., Bosco, F. A., & Kepes, S. (2020). How robust is our cumulative knowledge on turnover? *Journal of Business and Psychology*. https://doi.org/10.1007/s10869-020-09687-3

***Ford, M. T., Heinen, B. A., & Langkamer, K. L. (2007). Work and family satisfaction and conflict: A meta-analysis of cross-domain relations. *Journal of Applied Psychology*, *92*(1), 57–80. https://doi.org/10.1037/0021-9010.92.1.57

*Gajendran, R. S., & Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences. *Journal of Applied Psychology*, *92*(6), 1524–1541. https://doi.org/10.1037/0021-9010.92.6.1524

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.3102/0013189X005010003

***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*(6), 1222–1243. https://doi.org/10.1037/a0037547

***Greco, L. M., Whitson, J. A., O'Boyle, E. H., Wang, C. S. & Kim, J. (2019). An eye for an eye? A meta-analysis of negative reciprocity in organizations. *Journal of Applied Psychology*, *104*(9), 1117–1143. https://doi.org/10.1037/apl0000396

***Greer, L., Jong, B., Schouten, M. & Dannals, J. (2018). Why and When Hierarchy Impacts Team Effectiveness: A Meta-Analytic Integration. *Journal of Applied Psychology*, *103*(6), 591–613. https://doi.org/10.1037/apl0000291

Greenberg, S. A. (2009). How citation distortions create unfounded authority: Analysis of a citation network. *BMJ*, *339*, Article b2680. https://doi.org/10.1136/bmj.b2680

**Harari, D., Swider, B. W., Steed, L. B. & Breidenthal, A. P. (2018). Is perfect good? A meta-analysis of perfectionism in the workplace. *Journal of Applied Psychology*, *103*(10), 1121–1144. https://doi.org/10.1037/apl0000324

*Harrison, D. A., Kravitz, D. A., Mayer, D. M., Leslie, L. M., & Lev-Arey, D. (2006). Understanding attitudes toward affirmative action programs in employment: Summary and meta-analysis of 35 years of research. *Journal of Applied Psychology*, *91*(5), 1013–1036. https://doi.org/10.1037/0021-9010.91.5.1013

Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H., & Short, J. (2017). Publication bias in strategic management research. *Journal of Management*, *43*(2), 400–425. https://doi.org/10.1177/0149206314535438

*Hartnell, C. A., Ou, A. Y., Kinicki, A. J., Choi, D. & Karam, E. P. (2019). A meta-analytic test of organizational culture's association with elements of an organization's system and its relative predictive validity on organizational outcomes. *Journal of Applied Psychology*, *104*(6), 832–850. https://doi.org/10.1037/apl0000380

*Hartnell, C. A., Ou, A. Y., & Kinicki, A. (2011). Organizational culture and organizational effectiveness: A meta-analytic investigation of the competing values framework's theoretical suppositions. *Journal of Applied Psychology*, *96*(4), 677–694. https://doi.org/10.1037/a0021987

Hattrup, K., Rock, J., & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology*, *82*(5), 656–664. https://doi.org/10.1037/0021-9010.82.5.656

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, *13*, Article e1002106. https://doi.org/10.1371/journal.pbio.1002106

**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*(3), 412–453. https://doi.org/10.1037/a0032380

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

**Heidemeier, H., & Moser, K. (2009). Self-other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology*, *94*(2), 353–370. https://doi.org/10.1037/0021-9010.94.2.353

**Hoffman, B. J., Kennedy, C. L., LoPilato, A. C., Monahan, E. L., & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, *100*(4), 1143–1168. https://doi.org/10.1037/a0038707

*Hom, P. W., Caranikas-Walker, F., Prussia, G. E., & Griffeth, R. W. (1992). A meta-analytical structural equations analysis of a model of employee turnover. *Journal of Applied Psychology*, *77*(6), 890–909. https://doi.org/10.1037/0021-9010.77.6.890

***Hong, Y., Liao, H., Hu, J., & Jiang, K. (2013). Missing link in the service profit chain: A meta-analytic review of the antecedents, consequences, and moderators of service climate. *Journal of Applied Psychology*, *98*(2), 237–267. https://doi.org/10.1037/a0031666

**Howard, M. C., Cogswell, J. E. & Smith, M. B. (2020). The antecedents and outcomes of workplace ostracism: A meta-analysis. *Journal of Applied Psychology*, *105*(6), 577–596. https://doi.org/10.1037/apl0000453

***Huang, J. L., Cropanzano, R., Li, A., Shao, P., Zhang, X.-a., & Li, Y. (2017). Employee conscientiousness, agreeableness, and supervisor justice rule compliance: A three-study investigation. *Journal of Applied Psychology*, *102*(11), 1564–1589. https://doi.org/10.1037/apl0000248

*Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: A meta-analytic investigation. *Journal of Applied Psychology*, *99*(1), 162–179. https://doi.org/10.1037/a0034285

*Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., Benishek, L. E., King, H. B., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. *Journal of Applied Psychology*, *101*(9), 1266–1304. https://doi.org/10.1037/apl0000120

**Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *Journal of Applied Psychology*, *94*(5), 1128–1145. https://doi.org/10.1037/a0015978

*Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: A meta-analytic summary and theoretical extension of the work design literature. *Journal of Applied Psychology*, *92*(5), 1332–1356. https://doi.org/10.1037/0021-9010.92.5.1332

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings* (1st ed.). SAGE Publications.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). SAGE Publications.

*Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology*, *85*(6), 869–879. https://doi.org/10.1037/0021-9010.85.6.869

Ioannidis, J. P. A. (2005a). Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, *294*(2), 218–228. https://doi.org/10.1001/jama.294.2.218

Ioannidis, J. P. A. (2005b). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*(3), 245–253. https://doi.org/10.1177/1740774507079441

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, *3*(1), 109–135.

*Jenkins, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology*, *83*(5), 777–787. https://doi.org/10.1037/0021-9010.83.5.777

***Jiang, K., Liu, D., McKay, P. F., Lee, T. W., & Mitchell, T. R. (2012). When and how is job embeddedness predictive of turnover? A meta-analytic investigation. *Journal of Applied Psychology*, *97*(5), 1077–1096. https://doi.org/10.1037/a0028610

***Joseph, D. L., Jin, J., Newman, D. A., & O'Boyle, E. H. (2015). Why does self-reported emotional intelligence predict job performance? A meta-analytic investigation of mixed EI. *Journal of Applied Psychology*, *100*(2), 298–342. https://doi.org/10.1037/a0037681

*Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, *89*(3), 542–552. https://doi.org/10.1037/0021-9010.89.3.542

**Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875–925. https://doi.org/10.1037/a0033901

*Kammeyer-Mueller, J. D., Judge, T. A., & Scott, B. A. (2009). The role of core self-evaluations in the coping process. *Journal of Applied Psychology*, *94*(1), 177–195. https://doi.org/10.1037/a0013214

**Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, *93*(1), 59–69. https://doi.org/10.1037/0021-9010.93.1.59

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, *15*(4), 624–662. https://doi.org/10.1177/1094428112452760

Kepes, S., Banks, G. C., & Oh, I. S. (2014). Avoiding bias in publication bias research: The value of “null” findings. *Journal of Business and Psychology*, *29*(2), 183–203. https://doi.org/10.1007/s10869-012-9279-0

Kepes, S., & McDaniel, M. A. (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PLoS ONE*, *10*(10), Article e0141468. https://doi.org/10.1371/journal.pone.0141468

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*(2), 123–143. https://doi.org/10.1007/s10869-013-9300-2

Kepes, S., & Thomas, M. A. (2018). Assessing the robustness of meta-analytic results in information systems: Publication bias and outliers. *European Journal of Information Systems*, *27*(1), 90–123. https://doi.org/10.1080/0960085X.2017.1390188

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kinicki, A. J., Mckee-Ryan, F. M., Schriesheim, C. A., & Carson, K. P. (2002). Assessing the construct validity of the job descriptive index: A review and meta-analysis. *The Journal of Applied Psychology*, *87*(1), 14–32. https://doi.org/10.1037/0021-9010.87.1.14

**Kleingeld, A., van Mierlo, H., & Arends, L. (2011). The effect of goal setting on group performance: A meta-analysis. *Journal of Applied Psychology*, *96*(6), 1289–1304. https://doi.org/10.1037/a0024315

***Knight, A. P., & Eisenkraft, N. (2015). Positive is usually good, negative is not always bad: The effects of group affect on social integration and task performance. *Journal of Applied Psychology*, *100*(4), 1214–1227. https://doi.org/10.1037/apl0000006

**Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*(1), 128–161. https://doi.org/10.1037/a0036734

***Kooij, D. T. A. M., Kanfer, R., Betts, M. & Rudolph, C. W. (2018). Future time perspective: A systematic review and meta-analysis. *Journal of Applied Psychology*, *103*(8), 867–893. https://doi.org/10.1037/apl0000306

Koslowsky, M., & Sagie, A. (1993). On the efficacy of credibility intervals as indicators of moderator effects in meta-analytic research. *Journal of Organizational Behavior, 14,* 695-699.

**Koslowsky, M., Sagie, A., Krausz, M., & Singer, A. D. (1997). Correlates of employee lateness: Some theoretical considerations. *Journal of Applied Psychology*, *82*(1), 79–88. https://doi.org/10.1037/0021-9010.82.1.79

**Kossek, E. E., & Ozeki, C. (1998). Work-family conflict, policies, and the job-life satisfaction relationship: A review and directions for organizational behavior-human resources research. *Journal of Applied Psychology*, *83*(2), 139–149. https://doi.org/10.1037/0021-9010.83.2.139

*Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology*, *70*(1), 56–65. https://doi.org/10.1037/0021-9010.70.1.56

*Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*(6), 1060–1072. https://doi.org/10.1037/a0034156

*Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. *Journal of Applied Psychology*, *102*(12), 1686–1718. https://doi.org/10.1037/apl0000241

***Landay, K., Harms, P. D. & Credé, M. (2019). Shall we serve the dark lords? A meta-analytic review of psychopathy and leadership. *Journal of Applied Psychology*, *104*(1), 183–196. https://doi.org/10.1037/apl0000357

Landis, R. S., & Rogelberg, S. G. (2013). Our scholarly practices are derailing our progress: The importance of “nothing” in the Organizational Sciences. *Industrial and Organizational Psychology*, *6*(3), 299–302. https://doi.org/10.1111/iops.12054

***Lee, Y., Berry, C. M. & Gonzalez-Mulé, E. (2019). The importance of being humble: A meta-analysis and incremental validity analysis of the relationship between honesty-humility and job performance. *Journal of Applied Psychology*, *104*(12), 1535–1546. https://doi.org/10.1037/apl0000421

*Lee, K., Carswell, J. J., & Allen, N. J. (2000). A meta-analytic review of occupational commitment: Relations with person- and work-related variables. *Journal of Applied Psychology*, *85*(5), 799–811. https://doi.org/10.1037/0021-9010.85.5.799

***Leslie, L. M., Bono, J. E., Kim, Y. S. & Beaver, G. R. (2020). On melting pots and salad bowls: A meta-analysis of the effects of identity-blind and identity-conscious diversity ideologies. *Journal of Applied Psychology*, *105*(5), 453–471. https://doi.org/10.1037/apl0000446

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of doing research*. Harvard University Press.

Lin, L., Chu, H., & Hodges, J. S. (2017). Alternative measures of between-study heterogeneity in meta-analysis: Reducing the impact of outlying studies. *Biometrics*, *73*(1), 156–166. https://doi.org/10.1111/biom.12543

Lin, L., Chu, H., Murad, M. H., Hong, C., Qu, Z., Cole, S. R., & Chen, Y. (2018). Empirical comparison of publication bias tests in meta-analysis. *Journal of General Internal Medicine*, *33*(8), 1260–1267. https://doi.org/10.1007/s11606-018-4425-7

***Litwiller, B., Snyder, L. A., Taylor, W. D., & Steele, L. M. (2017). The relationship between sleep and work: A meta-analysis. *Journal of Applied Psychology*, *102*(4), 682–699. https://doi.org/10.1037/apl0000169

*Loher, B. T., Noe, R. A., Moeller, N. L., & Fitzgerald, M. P. (1985). A meta-analysis of the relation of job characteristics to job satisfaction. *Journal of Applied Psychology*, *70*(2), 280–289. https://doi.org/10.1037/0021-9010.70.2.280

***Longmire, N. H. & Harrison, D. A. (2018). Seeing their side versus feeling their pain: Differential consequences of perspective-taking and empathy at work. *Journal of Applied Psychology*, *103*(8), 894–915. https://doi.org/10.1037/apl0000307

*Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*(3), 280–296. https://doi.org/10.1037//0021-9010.67.3.280

**Martocchio, J. J., & O'Leary, A. M. (1989). Sex differences in occupational stress: A meta-analytic review. *Journal of Applied Psychology*, *74*(3), 495–501. https://doi.org/10.1037/0021-9010.74.3.495

**Mathieu, J. E., Kukenberger, M. R., D'Innocenzo, L., & Reilly, G. (2015). Modeling reciprocal team cohesion-performance relationships, as impacted by shared leadership and members' competence. *Journal of Applied Psychology*, *100*(3), 713–734. https://doi.org/10.1037/a0038898

***McCord, M. A., Joseph, D. L., Dhanani, L. Y. & Beus, J. M. (2018). A meta-analysis of sex and race differences in perceived workplace mistreatment. *Journal of Applied Psychology*, *103*(2), 137–163. https://doi.org/10.1037/apl0000250

McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, *59*(4), 927–953. https://doi.org/10.1111/j.1744-6570.2006.00059.x

*McEvoy, G. M., & Cascio, W. F. (1985). Strategies for reducing employee turnover: A meta-analysis. *Journal of Applied Psychology*, *70*(2), 342–353. https://doi.org/10.1037/0021-9010.70.2.342

*McKee-Ryan, F. M., Song, Z., Wanberg, C. R., & Kinicki, A. J. (2005). Psychological and physical well-being during unemployment: A meta-analytic study. *Journal of Applied Psychology*, *90*(1), 53–76. https://doi.org/10.1037/0021-9010.90.1.53

***McNatt, D. B. (2000). Ancient pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, *85*(2), 314–322. https://doi.org/10.1037/0021-9010.85.2.314

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Miller, K. D., & Tsang, E. W. K. (2011). Testing management theories: Critical realist philosophy and research methods. *Strategic Management Journal*, *32*(2), 139–158. https://doi.org/10.1002/smj.868

***Mitra, A., Jenkins, G. D., & Gupta, N. (1992). A meta-analytic review of the relationship between absence and turnover. *Journal of Applied Psychology*, *77*(6), 879–889. https://doi.org/10.1037/0021-9010.77.6.879

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

***Morris, S. B., Daisley, R. L., Wheeler, M., & Boyer, P. (2015). A meta-analysis of the relationship between individual assessments and job performance. *Journal of Applied Psychology*, *100*(1), 5–20. https://doi.org/10.1037/a0036938

Muggeo, V. M. R. (2016). Testing with a nuisance parameter present only under the alternative: A score-based approach with application to segmented modelling. *Journal of Statistical Computation and Simulation*, *86*(15), 3059–3067. https://doi.org/10.1080/00949655.2016.1149855

***Nahrgang, J. D., Morgeson, F. P., & Hofmann, D. A. (2011). Safety at work: A meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. *Journal of Applied Psychology*, *96*(1), 71–94. https://doi.org/10.1037/a0021484

**Narby, D. J., Cutler, B. L., & Moran, G. (1993). A meta-analysis of the association between authoritarianism and jurors' perceptions of defendant culpability. *Journal of Applied Psychology*, *78*(1), 34–42. https://doi.org/10.1037/0021-9010.78.1.34

***Ng, T. W. H., & Feldman, D. C. (2015). Ethical leadership: Meta-analytic evidence of criterion-related and incremental validity. *Journal of Applied Psychology*, *100*(3), 948–965. https://doi.org/10.1037/a0038246

**Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*(6), 1314–1334. https://doi.org/10.1037/a0012702

***Nohe, C., Meier, L. L., Sonntag, K., & Michel, A. (2015). The chicken or the egg? A meta-analysis of panel studies of the relationship between work-family conflict and strain. *Journal of Applied Psychology*, *100*(2), 522–536. https://doi.org/10.1037/a0038012

Nuijten, M. B., Van Assen, M. A. L. M., Augusteijn, H. E. M., Crompvoets, E. A. V., & Wicherts, J. M. (2020). Effect sizes, power, and biases in intelligence research: A meta-meta-analysis. *Journal of Intelligence*, *8*(4), 1–24. https://doi.org/10.3390/jintelligence8040036

O'Boyle, E. H., Banks, G. C., Carter, K., Walter, S., & Yuan, Z. (2019). A 20-year review of outcome reporting bias in moderated multiple regression. *Journal of Business and Psychology*, *34*(1), 19–37. https://doi.org/10.1007/s10869-018-9539-8

*O'Boyle, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *Journal of Applied Psychology*, *97*(3), 557–579. https://doi.org/10.1037/a0025679

O'Boyle, E. H., Rutherford, M. W., & Banks, G. C. (2014). Publication bias in entrepreneurship research: An examination of dominant relations to performance. *Journal of Business Venturing*, *29*(6), 773–784. https://doi.org/10.1016/j.jbusvent.2013.10.001

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

**Park, T. Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology*, *98*(2), 268–309. https://doi.org/10.1037/a0030723

***Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology*, *99*(6), 1129–1145. https://doi.org/10.1037/a0036751

*Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, *92*(1), 128–150. https://doi.org/10.1037/0021-9010.92.1.128

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*, *295*(6), 676. https://doi.org/10.1001/jama.295.6.676

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010

Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience and Biobehavioral Reviews*, *57*, 411–432. https://doi.org/10.1016/j.neubiorev.2015.09.017

Pietschnig, J., Siegel, M., Eder, J. S. N., & Gittler, G. (2019). Effect declines are systematic, strong, and ubiquitous: A meta-meta-analysis of the decline effect in intelligence research. *Frontiers in Psychology*, *10*, 1–13. https://doi.org/10.3389/fpsyg.2019.02874

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect-Shmozart effect: A meta-analysis. *Intelligence*, *38*(3), 314–323. https://doi.org/10.1016/j.intell.2010.03.001

***Porter, C. M., Woo, S. E., Allen, D. G. & Keith, M. G. (2019). How do instrumental and expressive network positions relate to turnover? A meta-analytic investigation. *Journal of Applied Psychology*, *104*(4), 511–536. https://doi.org/10.1037/apl0000351

*Premack, S. L., & Wanous, J. P. (1985). A meta-analysis of realistic job preview experiments. *Journal of Applied Psychology*, *70*(4), 706–719. https://doi.org/10.1037/0021-9010.70.4.706

Protzko, J., & Schooler, J. W. (2017). Decline effects: Types, mechanisms, and personal reflections. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 85–107). John Wiley & Sons, Ltd.

***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*(6), 1011–1041. https://doi.org/10.1037/a0037712

Renkewitz, F., & Keiner, M. (2019). How to detect publication bias in psychological pesearch: A comparative evaluation of six statistical methods. *Zeitschrift Für Psychologie/Journal of Psychology*, *227*(4), 261–279. https://doi.org/10.1027/2151-2604/a000386

*Rhoades, L., & Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology*, *87*(4), 698–714. https://doi.org/10.1037/0021-9010.87.4.698

**Riketta, M. (2008). The causal relation between job attitudes and performance: A meta-analysis of panel studies. *Journal of Applied Psychology*, *93*(2), 472–481. https://doi.org/10.1037/0021-9010.93.2.472

**Robbins, J. M., Ford, M. T., & Tetrick, L. E. (2012). Perceived unfairness and employee health: A meta-analytic integration. *Journal of Applied Psychology*, *97*(2), 235–272. https://doi.org/10.1037/a0025408

**Robertson, I. T., & Downs, S. (1989). Work-sample tests of trainability: A meta-analysis. *Journal of Applied Psychology*, *74*(3), 402–410. https://doi.org/10.1037/0021-9010.74.3.402

***Rockstuhl, T., Dulebohn, J. H., Ang, S., & Shore, L. M. (2012). Leader-member exchange (LMX) and culture: A meta-analysis of correlates of LMX across 23 countries. *Journal of Applied Psychology*, *97*(6), 1097–1130. https://doi.org/10.1037/a0029978

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rothstein, H. R., & Bushman, B. J. (2012). Publication bias in psychological science: Comment on Ferguson and Brannick (2012). *Psychological Methods*, *17*(1), 129–136. https://doi.org/10.1037/a0027128

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. https://doi.org/10.1002/0470870168

Ryan, A. M., & Nguyen, H.-H. D. (2017). Publication bias and stereotype threat research: A reply to Zigerell. *Journal of Applied Psychology, 102*(8), 1169–1177. https://doi.org/10.1037/apl0000242

***Sackett, P. R., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology*, *102*(10), 1435–1447. https://doi.org/10.1037/apl0000236

*Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, *82*(1), 30–43. https://doi.org/10.1037/0021-9010.82.1.30

Schalken, N., & Rietbergen, C. (2017). The reporting quality of systematic reviews and meta-analyses in Industrial and Organizational Psychology: A systematic review. *Frontiers in Psychology*, *8*, 1–12. https://doi.org/10.3389/fpsyg.2017.01395

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. SAGE Publications, Ltd. https://doi.org/10.4135/9781483398105

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed vs. random models in meta-analysis: Model properties and comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*, 97–128. https://doi.org/10.1348/000711007X255327

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, *470*(7335), 437. https://doi.org/10.1038/470437a

**Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology*, *104*(12), 1514–1534. https://doi.org/10.1037/apl0000420

**Shockley, K. M., Shen, W., DeNunzio, M. M., Arvan, M. L., & Knudsen, E. A. (2017). Disentangling the relationship between gender and work-family conflict: An integration of theoretical perspectives using meta-analytic methods. *Journal of Applied Psychology*, *102*(12), 1601–1635. https://doi.org/10.1037/apl0000246

Siegel, M., Eder, J. S. N., Gittler, G., Voracek, M., Wicherts, J. M., & Pietschnig, J. (2019). More than meets the eye: The prevalence of publication bias in Industrial and Organizational Psychology. *International Convention of Psychological Science (ICPS), 07.03. –09.03.2019, Paris, France*.

*Sin, H.-P., Nahrgang, J. D., & Morgeson, F. P. (2009). Understanding why they don't see eye to eye: An examination of leader-member exchange (LMX) agreement. *Journal of Applied Psychology*, *94*(4), 1048–1057. https://doi.org/10.1037/a0014827

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Simmons, J., & Nelson, L. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking. *Journal of Experimental Psychology: General*, *144*(6), 1146–1152.

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science*, *8*(5), 581–591. https://doi.org/10.1177/1948550617693062

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, *144*(12), 1325–1346. https://doi.org/10.1037/bul0000169

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

***Steel, R. P., & Ovalle, N. K. (1984). A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover. *Journal of Applied Psychology*, *69*(4), 673–686. https://doi.org/10.1037/0021-9010.69.4.673

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 76–98). John Wiley & Sons, Ltd.

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). John Wiley & Sons, Ltd.

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Online)*, *343*(7818), 1–8. https://doi.org/10.1136/bmj.d4002

**Stewart, W. H., & Roth, P. L. (2001). Risk propensity differences between entrepreneurs and managers: A meta-analytic review. *Journal of Applied Psychology*, *86*(1), 145–153. https://doi.org/10.1037/0021-9010.86.1.145

*Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Transfer of management training from alternative perspectives. *Journal of Applied Psychology*, *94*(1), 104–121. https://doi.org/10.1037/a0013006

van Aert, R. C. M., & van Assen, M. A. L. M. (2018). *P-uniform**. MetaArXiv. https://doi.org/10.31222/osf.io/zqjr9

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p-values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science*, *11*(5), 713–729. https://doi.org/10.1177/1745691616650874

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLoS ONE*, *14*(4), 1–32. https://doi.org/10.1371/journal.pone.0215052

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293–309. https://doi.org/10.1037/met0000025

van der Steen, J. T., van den Bogert, C. A., van Soest-Poortvliet, M. C., Fazeli Farsani, S., Otten, R. H. J., ter Riet, G., & Bouter, L. M. (2018). Determinants of selective reporting: A taxonomy based on content analysis of a random selection of the literature. *PLoS ONE*, *13*(2), e0188247. https://doi.org/10.1371/journal.pone.0188247

**Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, *96*(6), 1167–1194. https://doi.org/10.1037/a0024343

*Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, *97*(3), 499–530. https://doi.org/10.1037/a0021196

Vevea, J. L., Coburn, K. M., & Sutton, A. J. (2019). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 383–429). Russell Sage Foundation.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125. https://doi.org/10.1002/jrsm.11

**Vinchur, A. J., Schippmann, J. S., Switzer, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, *83*(4), 586–597. https://doi.org/10.1037//0021-9010.83.4.586

Voracek, M. (2014). No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D: 4D): A comprehensive meta-analysis and critical evaluation of research. *Evolution and Human Behavior*, *35*(5), 430–437. https://doi.org/10.1016/j.evolhumbehav.2014.05.009

***Wang, G., Van Iddekinge, C. H., Zhang, L. & Bishoff, J. (2019). Meta-analytic and primary investigations of the role of followers in ratings of leadership behavior in organizations. *Journal of Applied Psychology*, *104*(1), 70–106. https://doi.org/10.1037/apl0000345

**Wang, D., Waldman, D. A., & Zhang, Z. (2014). A meta-analysis of shared leadership and team effectiveness. *Journal of Applied Psychology*, *99*(2), 181–198. https://doi.org/10.1037/a0034531

***Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology*, *82*(2), 247–252. https://doi.org/10.1037/0021-9010.82.2.247

**Whitman, D. S., Caleo, S., Carpenter, N. C., Horner, M. T., & Bernerth, J. B. (2012). Fairness at the collective level: A meta-analytic examination of the consequences and boundary conditions of organizational justice climate. *Journal of Applied Psychology*, *97*(4), 776–791. https://doi.org/10.1037/a0028021

Wicherts, J. M. (2017). The weak spots in contemporary science (and how to fix them). *Animals*, *7*(12), 1–19. https://doi.org/10.3390/ani7120090

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*, 1–12. https://doi.org/10.3389/fpsyg.2016.01832

**Wright, T. A., & Bonett, D. G. (2002). The moderating effects of employee tenure on the relation between organizational commitment and job performance: A meta-analysis. *Journal of Applied Psychology*, *87*(6), 1183–1190. https://doi.org/10.1037/0021-9010.87.6.1183

***Zhang, Y., Zhang, Y., Ng, T. W. H. & Lam, S. S. K. (2019). Promotion- and prevention-focused coping: A meta-analytic examination of regulatory strategies in the work stress process. *Journal of Applied Psychology*, *104*(10), 1296–1323. https://doi.org/10.1037/apl0000404

*Zhao, H., & Seibert, S. E. (2006). The Big Five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology*, *91*(2), 259–271. https://doi.org/10.1037/0021-9010.91.2.259

Zigerell, L. J. (2017). Potential publication bias in the stereotype threat literature: Comment on Nguyen and Ryan (2008). *Journal of Applied Psychology*, *102*(8), 1159–1168. https://doi.org/10.1037/apl0000188

Footnotes

[1] We extracted the primary data as reported in the meta-analysis or as provided by the authors. The only exception to this rule is the data reported in Nguyen and Ryan (2008). In reply to a reanalysis of this meta-analysis (Zigerell, 2017), Ryan and Nguyen (2017) provided a corrected version of their 2008 dataset that we used for analysis instead of the data that had been originally reported. Consequently, we based our analyses on the corrected dataset as provided by Ryan and Nguyen (2017), but coded this dataset as belonging to Nguyen and Ryan (2008). Both Ryan and Ngyuen (2017) and Zigerell (2017) were not included in our meta-meta-analysis, as they do not report an original meta-analysis. Note that data from Shewach et al. (2019), also focusing on stereotype threat, is included because of the moderate proportion of overlapping samples (38%) with Nguyen and Ryan (2008).

[2] In order to rule out bias introduced by differing investigated time periods, we recalculated our prevalence estimate and time trend for the period 1995–2008 and found a prevalence rate of 27.7% as well as a negative but non-significant time trend (r_{pb} = -.09, df = 34, p = .621).

Table 1

Methodological Specifications and Thresholds for Bias Indication by Publication Bias Detection Method

Method	Methodological specifications	Threshold of bias indication
Trim-and-Fill (Duval & Tweedie, 2000a, 2000b)	Random-effects trim-and-fill estimation (estimator: L0), side of imputed studies is specified according to sign of summary effect (i.e., studies are imputed on the left side for positive summary effects and on the right side for negative summary effects).	Difference between the meta-analytic summary effect and the adjusted estimate exceeds 20% of the meta-analytic summary effect (in either direction).
Begg & Mazumdar's rank correlation test (Begg & Mazumdar, 1994)	Rank correlation test (based on Kendall's τ^2) as described by Begg & Mazumdar (1994).	$p < .10$
Sterne & Egger's regression approach (Sterne & Egger, 2005)	Regression test as described in Sterne & Egger (2005), based on a random-effects model (estimator: maximum likelihood) and the standard error as predictor.	$p < .10$

Method	Methodological specifications	Threshold of bias indication
PET-PEESE (Stanley & Doucouliagos, 2014)	PET-PEESE procedure as described in Stanley & Doucouliagos (2014) based on a random-effects model (estimator: maximum-likelihood) and the standard error (PET) and the variance (PEESE) as predictors in the respective models.	Difference between the meta-analytic summary effect and the adjusted estimate exceeds 20% of the meta-analytic summary effect (in either direction). We interpreted PEESE estimates when the coefficient of the intercept was significant at $p < .10$ and PET estimates in all other cases (Stanley, 2017).
p-curve (Simonsohn et al., 2014)	*p*-curve procedure as described in Simonsohn et al. (2015), Z-values for continuous tests are obtained using Stouffer's method. Only effect sizes in the same direction as the summary effect are included.	*p*-value of half *p*-curve > .05 or *p*-values of both half and full *p*-curves > .10, corresponding to the combination test as described in Simonsohn et al. (2015).
p-uniform (van Assen et al., 2015)	*p*-uniform analysis (estimator: P [Irwin-Hall distribution]) as described in van Assen et al. (2015) using correlation effect sizes and sample sizes as input. Side is specified according to sign of summary effect (i.e., studies are imputed on the left	$p < .10$ for one-sided test for publication bias.

Method	Methodological specifications	Threshold of bias indication
	side for positive summary effects and on the right side for negative summary effects).	
p-uniform* (van Aert & van Assen, 2018)	*p*-uniform* as described in van Aert & van Assen (2018), using maximum-likelihood estimation and correlation effect sizes and sample sizes as input.	$p < .10$ for one-sided test for publication bias.
Test of Excess Significance (Ioannidis & Trikalinos, 2007)	Test of Excess Significance as described in Ioannidis & Trikalinos (2007).	$p < .10$ & observed number of significant effect sizes (O) exceeds expected number of significant effect sizes (E).
Selection Models (Vevea & Woods, 2005)	Moderate one-tailed selection model using weights and cut-offs as described in Vevea & Woods (2005).	Difference between the meta-analytic summary effect and the adjusted estimate exceeds 20% of the meta-analytic summary effect (in either direction).

Table 2

Descriptive Meta-Analytic and Meta-Meta-Analytic Statistics for All Sets of Meta-Analyses

Variable	Full set	Sets with primary data ($k > 9$)		Sets with homogenous primary data (lowest quartile of τ^2-range and $k > 9$)	
		Full	Published	Full	Published
Number of meta-analyses	128	87	84	22	21
Range of publication years (meta-analyses)	1982–2020	1984–2020	1984–2020	1992–2019	1992–2018
Range of covered timespan within meta-analyses	5–76 (30.94)	8–76 (32.14)	8–76 (32.68)	12–75 (35.91)	12–60 (34.19)
Hunter & Schmidt / Hedges & Olkin	117/11	78/9	75/9	21/1	20/1
Number of reported effect sizes	7,263 (56.74)	5,273 (60.61)	5,231 (62.27)	1,593 (72.41)	1,523 (72.52)
Number of reported participants	3,209,663 (25,075.49)	2,810,908 (32,309.29)	2,801,218 (33,347.83)	1,540,511 (70,023.23)	1,525,425 (72,639.29)
Datasets used for recalculation					
Number of effect sizes (recalculated)		4,988 (57.33)	3,705 (44.11)	1,452 (66.00)	896 (42.67)

Variable	Full set	Sets with primary data (k > 9)		Sets with homogenous primary data (lowest quartile of τ^2-range and k > 9)	
		Full	Published	Full	Published
Number of participants in recalculations		2,801,851 (32,205.18)	1,415,519.4 (16,851.42)	1,536,793 (69,854.24)	603,428.2 (28,734.68)
Meta-meta-analytic summary effect (abs.)		0.26 (0.02)	0.26 (0.02)	0.14 (0.02)	0.16 (0.02)
I^2 (weighted/unweighted)		81.65/77.24	81.27/77.09	79.82/61.21	76.99/59.67

Note. Excepting the meta-meta-analytic summary effect (standard error), values in parentheses refer to mean values. Mean I^2-value weighted by number of effect sizes.

Table 3

Frequency and Time Trends of Publication Bias Investigation, Overall and by Method

	Overall	1982–2000	2001–2010	2011–2020	Time trend
Number of meta-analyses investigating bias	53/128 (.41)	8/25 (.32)	9/33 (.27)	36/70 (.51)	r_{pb} = .19 [.02; .35]
Mean / median / range of used methods	1.74, 1, 1–7	1, 1, 1–1	1.22, 1, 1–3	2.03, 2, 1–7	r_s = .33 [.16; .47] (overall) r_s = .53 [.31; .70] (bias inv.)
			Publication bias detection methods		
Direct comparison	24 (.19/.45)	—	5 (.15/.56)	19 (.27/.53)	r_{pb} = .24 [-.03; .48]
Failsafe *N*	20 (.16/.38)	8 (.32/1)	4 (.12/.44)	8 (.11/.22)	r_{pb} = -.54 [-.71; -.31]
Funnel plot	15 (.12/.28)	—	1 (.03/.11)	14 (.20/.39)	r_{pb} = .42 [.17; .62]
Trim-and-fill	15 (.12/.28)	—	—	15 (.21/.42)	r_{pb} = .46 [.22; .65]
Sterne & Egger's regression	4 (.03/.08)	—	—	4 (.06/.11)	r_{pb} = .23 [-.04; .47]
Begg & Mazumdar's rank test	3 (.02/.06)	—	—	3 (.04/.08)	r_{pb} = .20 [-.08; .45]
Cumulative meta-analysis	3 (.02/.06)	—	—	3 (.04/.08)	r_{pb} = .17 [-0.10;.42]
TES/*p*-curve/Selection Model	1 (.01/.02)	—	—	1 (.01/.03)	—

Note. r_{pb} = point-biserial correlation, r_s = Spearman's rank correlation, TES = Test of Excess Significance.

N = 53 for all time trends by method. Numbers in parentheses indicate percentages (by method: overall/meta-analyses that reported results of bias analyses). Numbers in brackets indicate 95% CIs. PET-PEESE and *p*-uniform were not applied in any meta-analysis and therefore omitted from the table.

Table 4

Number of Meta-Analyses Indicative of Publication Bias by Method and Subset

Method	Full set	Published set	Full set with homogenous studies (lowest quartile of τ^2-range) only	Published set with homogenous studies (lowest quartile of τ^2-range) only
PET-PEESE	39/87 (.45)	43/84 (.51)	12/22 (.55)	13/21 (.62)
Selection Model	29/84[a] (.35)	26/82[a] (.32)	9/19[a] (.47)	5/19[a] (.26)
Sterne & Egger's regression	22/87 (.25)	18/84 (.21)	5/22 (.23)	7/21 (.33)
Begg & Mazumdar's rank test	19/87 (.22)	14/84 (.17)	6/22 (.27)	4/21 (.19)
Trim-and-fill	18/87 (.21)	17/84 (.20)	5/22 (.23)	6/21 (.29)
TES	14/87 (.16)	13/84 (.15)	3/22 (.14)	3/21 (.14)
p-uniform*	6/86[a] (.07)	3/84 (.04)	1/22 (.05)	0/21 (0)
p-uniform	3/87 (.03)	3/84 (.04)	0/22 (0)	1/21 (.05)
p-curve	0/87 (0)	1/84 (.01)	0/22 (0)	1/21 (.05)

Note. TES = Test of Excess Significance. k/K = Number of times that the method indicated bias / number of meta-analyses in which we used this method. Number in parentheses indicate percentages.

[a] Differing overall number due to non-convergence.

Table 5

Number of Bias Detection Methods Indicative of Bias Within Meta-Analyses and Indication of Bias According to BRE/MRE

Set	Number of methods indicative of bias								BRE/MRE	
	k	0	1	2	3	4	5	6	negligible	non-negligible
Full	87	27	18	13	19	4	3	3	26	61
Full – homogenous (lowest quartile of τ^2-range)	22	5	4	4	7	2	—	—	5	17
Published	84	32	15	8	18	4	5	2	26	58
Published – homogenous (lowest quartile of τ^2-range)	21	6	5	3	2	2	3	—	5	16

Note. *k* = Number of reanalyzed meta-analytic datasets per meta-meta-analytic set. Cell entries indicate number of meta-analyses. BRE/MRE = baseline range estimate/maximum range estimate.

Table 6

Results of Final Theory-guided Hierarchical Stepwise Unweighted Logistic Meta-Meta-Regression Models Predicting Bias Indication According to the Combined BRE/MRE-Measure (0 = Negligible Bias; 1 = Non-Negligible Bias) from Meta-Analytic and Primary Study Characteristics and Using Either I^2 or τ^2 as Predictors in Step 3

Coefficient	Full set		Published set	
	ß (*SE*)	*p*	ß (*SE*)	*p*
	Step 3a: I^2			
Intercept	0.95 (0.26)	< .001	0.95 (0.27)	< .001
Meta-analytic summary effect	-0.94 (0.34)	.006	-1.09 (0.37)	.003
Initial effect	0.62 (0.35)	.074	0.98 (0.40)	.014
Number of effect sizes	> -0.01 (0.26)	.995	-0.08 (0.26)	.766
I^2	-0.02 (0.27)	.930	-0.18 (0.28)	.527
Pseudo-R^2: HL/CS/N	.090/.104/.148		.118/.135/.191	
AIC	106.541		101.730	
Step 1 vs. Step 3a	$\chi^2(3) = 3.72$, $p = .293$		$\chi^2(3) = 8.10$, $p = .044$	
Step 2 vs. Step 3a	$\chi^2(2) = 0.01$, $p = .996$		$\chi^2(2) = 0.65$, $p = .723$	
	Step 3b: τ^2			
Intercept	1.14 (0.30)	< .001	1.06 (0.30)	< .001
Meta-analytic summary effect	-1.65 (0.47)	< .001	-1.25 (0.41)	.002
Initial effect	0.72 (0.41)	.075	0.70 (0.45)	.115

Coefficient	Full set		Published set	
	ß (*SE*)	*p*	ß (*SE*)	*p*
Number of effect sizes	-0.03 (0.27)	.912	-0.20 (0.25)	.422
τ^2	1.17 (0.42)	.005	0.82 (0.41)	.044
Pseudo-R^2: HL/CS/NG	.199/.216/.306		.165/.185/.261	
AIC	94.955		96.763	
Step 1 vs. Step 3b	$\chi^2(3) = 15.31$, $p = .002$		$\chi^2(3) = 13.07$, $p = .004$	
Step 2 vs. Step 3b	$\chi^2(2) = 11.59$, $p = .003$		$\chi^2(2) = 5.62$, $p = .060$	

Note. ß = standardized regression coefficient (log-scale). HL = Hosmer & Lemeshow, CS = Cox & Snell, NG = Nagelkerke. Full models are reported in Supplement Table 22.

Table 7

Results of Final Theory-guided Hierarchical Stepwise Unweighted Poisson Meta-Meta-Regression Models Predicting Number of Methods Indicative of Bias from Meta-Analytic and Primary Study Characteristics and Using Either I^2 or τ^2 as Predictors in Step 3

Coefficient	Full set		Published set	
	ß (*SE*)	*p*	ß (*SE*)	*p*
Step 3a: I^2				
Intercept	0.40 (0.09)	< .001	0.25 (0.11)	.019
Meta-analytic summary effect	-0.65 (0.12)	< .001	-0.91 (0.13)	< .001
Initial effect	0.24 (0.10)	.018	0.37 (0.11)	< .001
Number of effect sizes	0.12 (0.07)	.078	0.06 (0.07)	.353
I^2	0.09 (0.09)	.341	0.12 (0.10)	.241
Pseudo-R^2: HL/CS/N	.251/.361/.434		.352/.513/.590	
AIC	287.345		263.704	
Step 1 vs. Step 3a	$\chi^2(3) = 8.87$, $p = .031$		$\chi^2(3) = 13.59$, $p = .004$	
Step 2 vs. Step 3a	$\chi^2(2) = 4.95$, $p = .084$		$\chi^2(2) = 2.97$, $p = .227$	
Step 3b: τ^2				
Intercept	0.36 (0.10)	< .001	0.22 (0.11)	.041
Meta-analytic summary effect	-0.75 (0.12)	< .001	-0.92 (0.13)	< .001
Initial effect	0.23 (0.10)	.017	0.28 (0.11)	.012

Coefficient	Full set		Published set	
	ß (*SE*)	*p*	ß (*SE*)	*p*
Number of effect sizes	0.14 (0.07)	.039	0.09 (0.07)	.216
τ^2	0.33 (0.08)	< .001	0.24 (0.08)	.002
Pseudo-R^2: HL/CS/NG	.340/.455/.547		.393/.552/.635	
AIC	273.541		256.658	
Step 1 vs. Step 3b	$\chi^2(3) = 22.68$, $p < .001$		$\chi^2(3) = 20.63$, $p < .001$	
Step 2 vs. Step 3b	$\chi^2(2) = 18.76$, $p < .001$		$\chi^2(2) = 10.01$, $p = .007$	

Note. ß = standardized regression coefficient. HL = Hosmer & Lemeshow, CS = Cox & Snell, NG = Nagelkerke. Full models are reported in Supplementary Table 23.

Figure Captions

Figure 1

Overview of Analyses on the Meta-Analytic and the Meta-Meta-Analytic Level

Figure note: [a] adjusted estimates included in BRE/MRE analysis. [b] analyses that were also grouped by topic. [c] analyses that were also conducted using only meta-analytic datasets with $I^2 < 25\%$, [d] analyses that were done on meta-analytic sets without influential cases. [e] analyses that were conducted on unreliability-corrected and corresponding observed sets of effect sizes.

Figure 2

Flow Chart of Study Selection Process

Figure 3

Time Trends of Publication Bias Detection Prevalence by Method and Overall

Figure 4

Time Trend of Number of Publication Bias Detection Methods