Central questions about meta-analyses in psychological research: An annotated reading list

Susanne Buecker¹ · Johannes Stricker² · Michael Schneider²

Accepted: 1 June 2021 / Published online: 22 June 2021
© The Author(s) 2021

Abstract

A large amount of literature is available for researchers who are interested in performing meta-analyses in psychology. However, due to a large number of available sources and meta-analytic approaches, it can be difficult to get started with a meta-analysis when prior experiences are limited. In this annotated reading list, we provide an overview of and comment on 12 recommended sources that address the most relevant questions for conducting and presenting meta-analyses in psychology. Additionally, we point to various further readings and software packages that address more specific meta-analytic topics. With this guide, we aim to provide a starting point for researchers who wish to conduct a meta-analysis and for reviewers and editors who evaluate the quality of manuscripts presenting meta-analytic findings.

Keywords Meta-analysis · Systematic review · Reading list

Background

Meta-analysis has become an important and widely used technique to synthesize research findings in psychology. Any researcher who has published a handful of meta-analyses will regularly receive questions from colleagues that are interested in conducting a meta-analysis. Depending on the colleagues' prior experience, their questions vary in complexity ranging from basics ("Where do I start with my meta-analysis?") to specifics (e.g., "Which methods are most appropriate for assessing publication bias?"). To date, there are several excellent and comprehensible introductions to conducting a meta-analysis in the form of introductory articles (e.g., Cheung & Vijayakumar, 2016) or books (e.g., Borenstein et al., 2009b; Cooper et al., 2009). However, due to the wide range of meta-analytical methods and procedures, none of the existing publications covers all aspects relevant to conducting a meta-analysis. In this manuscript, we present an annotated reading list covering the most relevant aspects for researchers aiming to

conduct a meta-analysis in psychology. This annotated reading list may also serve as a guide for reviewers and editors who evaluate meta-analytic studies. This article is structured so that the annotated literature refers to increasingly specific questions. In addition to the 12 sources that we discuss in more detail, we provide further reading recommendations containing additional information on the respective topic. We begin by describing what a meta-analysis is and what general best-practice recommendations for meta-analyses are available before introducing articles about the different phases of conducting a meta-analysis: literature search, coding and transforming effect sizes, artifact correction, statistical analyses, and quantifying heterogeneity and publication bias. We close by summarizing two articles about reporting standards for meta-analyses and open science principles.

Introduction: What Is a Meta-Analysis?

Source Field, A. P. (2005). Meta-analysis. In J. Miles & P. Gilbert (eds.), *A handbook of research methods in clinical and health psychology* (pp. 295–308). Oxford University Press. <https://doi.org/10.1093/med:psych/9780198527565.001.0001>

Field (2005) provides a "whistle-stop tour" (p. 307), introducing the most important elements of meta-analysis. He begins with a brief comparison of meta-analyses and discursive literature reviews, highlighting the advantages of the former.

Susanne Buecker and Johannes Stricker shared first-authorship.

✉ Susanne Buecker
susanne.buecker@rub.de

¹ Psychology Department, Ruhr University Bochum, Universitätsstr. 150, 44801 Bochum, Germany

² Psychology Department, University of Trier, Trier, Germany

According to Field, the problem with qualitative reviews is that different researchers might come to different conclusions based on the same literature. In contrast, meta-analyses provide a more objective approach by quantitatively rather than qualitatively aggregating the available evidence.

Field (2005) processed with briefly introducing different effect size measures as the essential basis for any meta-analysis. This overview of effect size measures is helpful in understanding different forms of meta-analysis (e.g., correlational or group comparisons). The mathematical formulas provided in the chapter are less important at this point but help at a later stage in conducting meta-analyses (see section “How to compute or transform effect sizes?”).

In the subsequent section, Field (2005) provides an overview of the required steps in a meta-analysis, including the literature search, inclusion criteria, and the calculation of effect sizes. He also explains the critical differences between random-effects and fixed-effects meta-analysis with the aid of two insightful illustrations. Additionally, Field (2005) provides a primer of two widely-used meta-analytic techniques. Finally, Field (2005) primes for highly relevant problems with meta-analysis. These problems may be methodological or conceptual. Examples include publication bias (selective reporting of significant results that distort meta-analytic estimates (see section “Which methods are most appropriate for assessing publication bias?”), artifacts such as measurement unreliability (see section “How can the meta-analytic results be corrected for artifacts?”), misapplications of meta-analysis (random-effects meta-analyses are recommended if the aim is to gain generalizable knowledge, whereas fixed-effects models are not recommended if the aim is to gain generalizable knowledge; also see Hunter & Schmidt, 2000¹), and other errors in methods. On the conceptual level, Field (2005) discusses more general objections to meta-analysis (e.g., oversimplification).

Overall, Field (2005) provides a comprehensive and accessible introduction to meta-analysis as a means to statistically aggregate effect sizes from a large number of primary studies. His overview of the required steps in a meta-analysis simplifies the understanding of the literature recommended here.

Further reading. Borenstein et al. (2009a, c) provide a more detailed introduction to meta-analyses, covering the history and historical development of meta-analyses (Borenstein et al., 2009c) and the rationale for statistically aggregating effect sizes (Borenstein et al., 2009a). An alternative classical introduction to meta-analysis is provided by Lipsey and Wilson (2001). This work provides great conceptual insights and hands-on recommendations for conducting meta-analyses

but, naturally, misses newer developments in meta-analyses from the last 20 years. In a more complex and theoretical manuscript, Chan and Arvey (2012) discuss the value of meta-analyses in the larger context of the progress of science more generally. We recommend this text for researchers who aim to develop a deeper understanding of the role of research aggregation in advancing knowledge growth.

How to Conduct a Literature Search for a Meta-Analysis in Psychology?

Source Harari, M. B., Parola, H. R., Hartwell, C. J., & Riegelman, A. (2020). Literature searches in systematic reviews and meta-analyses: A review, evaluation, and recommendations. *Journal of Vocational Behavior*, 118, 103,377. <https://doi.org/10.1016/j.jvb.2020.103377>

A comprehensive and reproducible literature search is crucial to ensure that a meta-analysis covers (as far as possible) all relevant primary studies. Harari et al. (2020) provide concrete best practice recommendations for conducting a literature search for a meta-analysis in psychology. Importantly, the authors stress that these recommendations are flexible depending on the characteristics of the literature covered. In Study 1, the authors report a systematic review of 152 reviews in applied psychology, identifying the most commonly used literature search approaches. Table 1 in Harari et al. (2020) provides an overview of complementary search strategies (e.g., searching the references of included articles) that may be used in addition to database searches. Table 3 in Harari et al. (2020) displays the most commonly searched databases in applied psychology. In Study 2, Harari et al. (2020) empirically demonstrate the effects of different literature search strategies on meta-analytic effect sizes.

In their discussion section, Harari et al. (2020) thoughtfully discuss the pros and cons of different databases (e.g., Google Scholar), complementary search strategies, and transparent reporting of search strategies. Table 5 in Harari et al. (2020) provides a comprehensive summary of all recommendations for conducting and reporting a literature search. These recommendations are the following: a) use at least two databases, b) involve a librarian in identifying the most relevant databases, c) indicate the databases searched, not only the platform used to search (e.g., ProQuest), d) transparently report all search information and search variables, e) transparently report the procedures and criteria for manual literature searches, f) incorporate backward citation search (i.e., search references of studies included in previous reviews), g) transparently report the procedures and criteria for backward searches, h) consider conducting a forward search of well-known classic papers on the respective topic, i) attempt to include non-indexed (or non-published) studies, j) search non-academic databases and websites. We recommend that researchers conducting a

¹ Fixed-effects meta-analyses typically manifest a substantial Type I bias in significance tests for mean effect sizes and for moderator variables, whereas random-effects meta-analyses do not. However, see Doi et al. (2015) for arguments against using random-effects meta-analytic models.

literature search for a meta-analysis in psychology adhere to these recommendations. Yet, there are circumstances where there are good and justifiable reasons for not implementing a recommendation (e.g., consulting a librarian when the researchers know the relevant databases, search terms, and operators well).

Further reading. In a database search, researchers use so-called Boolean operators (or Boolean phrases) to connect search terms. For example, OR may be used to indicate that an article must contain at least one of two words to be relevant (e.g., “Extraversion OR Introversion”). Another Boolean operator may be used to truncate search terms (e.g., perfection* to cover perfectionistic, perfectionist, perfectionism, etc.). In his blog post, Calhoun (2013) comprehensively introduces the most relevant Boolean operators and search limiters (e.g., publication year) for PsycINFO. Please note, however, that Boolean operators differ between databases (Othman & Sahlawaty Halim, 2004). Adjusting the search string (i.e., the search terms connected with Boolean operators) is usually necessary when searching through different databases. Thus, meta-analysts need to acquaint themselves with the relevant operators for all included databases. These operators are usually explained on the databases’ websites.

What Are the Best Practices in Conducting and Reporting Systematic Reviews?

Source Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70, 747–770. <https://doi.org/10.1146/annurev-psych-010418-102803>

The basis for any quantitative meta-analysis is a systematic and replicable review of the relevant empirical literature. Reviews are sometimes called narrative reviews or qualitative reviews to distinguish them from quantitative meta-analyses. Systematic reviews are narrative reviews that are based on a standardized and, thus, replicable literature search. In many cases, a systematic review and a meta-analysis are jointly presented. Siddaway et al. (2019) provide a comprehensive overview of the best practices in conducting and reporting systematic reviews. The authors start their article by explaining the advantages of systematic reviews with regard to drawing robust conclusions, explaining inconsistencies in findings, and generating implications for theory and future research. Against the background of the replication crisis in psychology, Siddaway et al. (2019) place particular emphasis on the increased transparency and replicability of a systematic review in comparison to a mere narrative review. Siddaway et al. (2019) also provide useful guidance on when to conduct

a meta-analysis, a narrative review, or a meta-synthesis: Meta-analyses are recommended when relevant primary quantitative studies that report standardized effect sizes and that examine similar constructs are available. Narrative reviews may be preferred when the relevant primary studies differ widely in their methods and investigated constructs or relations so that researchers have to connect formerly independent lines of research (for guidance on narrative reviews, see Baumeister, 2013; Baumeister & Leary, 1997). Meta-syntheses aggregate qualitative research. The main part of the article by Siddaway et al. (2019) describes the main stages in conducting a systematic review, from formulating the research question to assessing the quality of included studies. Particularly the sections explaining the identification of relevant search terms, the formulation of inclusion criteria, the creation of record-keeping systems, article screening, and eligibility assessment are practically relevant for meta-analysts. In the section on how to present a systematic review, Siddaway et al. (2019) make suggestions on how to structure the introduction, method, results, and discussion sections of a systematic review. These recommendations are useful for meta-analysts, however, there are some specificities for meta-analysis (see section “How should a meta-analysis be reported?”). In sum, Siddaway et al. (2019) provide essential insights into the rationale for systematic reviews or meta-analyses. This source is particularly relevant for the decisions whether to conduct a meta-analysis, how to carry out the required systematic literature review, and how to present the rationale for it.

Further reading. The Cochrane Collaboration provides a detailed handbook for systematic reviews of interventions (Higgins & Cochrane Collaboration, 2020). Part 2 of this book provides detailed guidance for systematic reviews that are useful for systematic reviews of intervention studies, ranging from determining the research question to interpreting results. Parts 3 and 4 of this handbook deal with more specific topics, some of which are relevant for psychology (e.g., the inclusion of non-randomized studies in a systematic review of interventions). The Cochrane handbook (Higgins & Cochrane Collaboration, 2020) is available free of charge online (<https://training.cochrane.org/handbook/current>).

How Many Effect Sizes Are Needed for a Meta-Analysis?

Source Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need?: A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35, 215–247. <https://doi.org/10.3102/1076998609346961>

After identifying the relevant primary studies, meta-analysts are inevitably confronted with the question of how many effect sizes are required to conduct a meta-analysis.

Valentine et al. (2010) provide a primer for statistical power for meta-analysis. As in primary studies, low statistical power may lead to difficulties in interpreting non-significant results. In their manuscript, Valentine et al. (2010) first illustrate how statistical power in primary studies and meta-analysis both depend on the sample size, the estimated effect size, and Type 1 error. The authors then explain the information and computations required to conduct prospective statistical power analyses for random-effects and fixed-effects meta-analytic models, including specific statistical tests such as the test of homogeneity and moderation analyses. The authors further explain how to establish the minimum number of studies required to achieve sufficient statistical power (e.g., 80%; Cohen, 1988) in different scenarios.

Retrospective statistical power analyses indicate how non-significant meta-analytic results may be interpreted considering the smallest relevant effect size. Valentine et al. (2010) provide a practical suggestion for wording in cases where low statistical power prevents meaningful interpretation of non-significant results. Retrospective power analysis may also guide the decision whether a meta-analysis requires an update, which is particularly relevant to researchers aiming to investigate a question for which an older meta-analysis already exists.

Another problem associated with statistical power is that trivially small effects may reach statistical significance in meta-analyses because of large sample sizes. Conversely, in the case of non-significant meta-analytic results, large confidence intervals may indicate that there is insufficient evidence to conclude whether the null hypothesis may be rejected. Thus, Valentine et al. (2010) argue in favor of interpreting effect sizes and confidence intervals in meta-analysis rather than p -values.

The last section of the article by Valentine et al. (2010) demonstrates that statistical aggregation is generally the most appropriate strategy for summarizing quantitative findings, even when the statistical power is low. For example, establishing that there is currently insufficient evidence to draw reliable conclusions about a research question combined with specific recommendations for future studies may be an important conclusion of the *review* part of a meta-analytic review. However, there may be circumstances in which researchers choose not to statistically aggregate primary studies (e.g., in the case of a very low number of studies with varying characteristics). For these circumstances, Valentine et al. (2010) suggest methods to review the available evidence narratively, highlighting current limitations of the evidence base. The authors conclude that, in principle, only two studies are necessary to justify a meta-analysis. Thus, researchers should feel encouraged to conduct a meta-analysis, even when there are only a few primary studies. However, statistical power should be considered in the interpretation of meta-analytic results.

Further reading. Hedges and Pigott (2004) provide a detailed account of power analyses for moderators in a meta-analysis, concluding that moderation analysis may, in many cases, suffer from low statistical power. Moreover, Jackson and Turner (2017) retrospectively calculated the power of 1991 meta-analyses taken from the Cochrane Database of Systematic Reviews. They conclude that, in practice, five or more studies are needed to achieve statistical power in a random-effects meta-analysis that is greater than the studies that contribute to them.

How to Compute or Transform Effect Sizes?

Source Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18. <https://doi.org/10.1037/a0024338>

Various statistical estimates quantitatively describe the magnitude of an empirical effect. When conducting a meta-analysis, researchers often encounter the situation that different primary studies report different effect size estimates. In these situations, meta-analysts have to transform an effect size into a different effect size, or in some cases, calculate effect sizes from other statistical values (e.g., means and standard deviations). In their article, Fritz et al. (2012) explain the rationale for using standardized effect size estimates (rather than, e.g., unstandardized mean differences) in meta-analyses and provide an overview of the most frequently used effect size estimates. This overview is followed by a review of articles from the *Journal of Experimental Psychology: General*, analyzing the statistical analyses reported and the associated reporting of effect size estimates. The subsequent section on calculating effect sizes is highly relevant for researchers interested in conducting a meta-analysis. In this section, the authors comprehensively explain how to compute and transform the most widely used effect sizes. These include effect sizes specific to comparing two conditions (e.g., experimental groups; Cohen's d , Hedges's g , Glass's d or Δ and point biserial correlation r) and effect sizes based on the proportion of variability explained (η^2 , R^2 , adjusted R^2). Additionally, Fritz et al. (2012) guide the interpretation of effect sizes. The concluding recommendations by Fritz et al. (2012) for reporting and using effect sizes are directed at researchers conducting primary studies rather than meta-analysts. In sum, the article by Fritz et al. (2012) is recommended as an overview of the relevant effect sizes and their interrelations.

Further reading. A large proportion of meta-analyses investigates relations between two continuous variables (e.g., personality and academic success). Primary studies investigating such relations usually report Pearson's correlation coefficient. However, in some cases, only partial correlations or

regression coefficients from multiple regressions (i.e., effect sizes controlled for the influence of other variables) are available. In a relatively technical article, Aloe (2015) demonstrates that these estimates usually deviate from bivariate effect sizes (such as Pearson's r). We recommend this manuscript for advanced meta-analysts. In short, Aloe (2015) shows that standardized regression coefficients only equal bivariate correlation coefficients when there is not more than one predictor in a regression analysis or the hypothetical case that all predictors in multiple regression are entirely uncorrelated ($r = .00$). In other cases, partial correlations or regression coefficients from multiple regressions should not be included in meta-analyses of bivariate relations.

For some meta-analyses, correlational studies reporting Pearson's correlation coefficients and group-comparison studies reporting Cohen's d are relevant. Pustejovsky (2014) provides recommendations for converting from a standardized mean difference to a correlation coefficient (and to Fisher's z) under three types of study designs: extreme groups, dichotomization of a continuous variable, and controlled experiments. Moreover, he provides information on how the sampling variance of effect size statistics should be estimated in each of these cases. Finally, we recommend Del Re's (2013) useful package `compute.es` for the conversion of effect sizes in the statistical environment R (R Core Team, 2020).

How Can the Meta-Analytic Results be Corrected for Artifacts?

Source Schmidt, F. L., & Hunter, J. E. (2015a). Meta-analysis of correlations corrected individually for artifacts. In *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed., pp. 87–164). SAGE Publications. <https://doi.org/10.4135/9781483398105>

There are various sources of error (i.e., study artifacts) in primary studies that can attenuate the magnitude of meta-analytic correlation coefficients. Schmidt and Hunter (2015a) introduce eleven of these study artifacts: sampling error, measurement error in the dependent—and/or independent variable, dichotomization of a continuous dependent—and/or independent variable, range restriction or enhancement in the dependent—and/or the independent variable, imperfect construct validity in the dependent—and/or the independent variable, reporting error, and variance due to third variables that affect the meta-analytic relation. In their approach to correcting meta-analyses for study artifacts, Schmidt and Hunter (2015a) propose correcting each correlation included in a meta-analysis individually. After an introduction to correcting error and bias in meta-analysis, Schmidt and Hunter (2015a) discuss meta-analysis with correction for sampling error only. Whereas some parts of this section are relatively technical, the provided examples are remarkably

illustrative. Following these examples, Schmidt and Hunter (2015a) explain in more detail the correction for attenuation due to measurement error (i.e., unreliable measurement of the included constructs), highlighting the relevance of selecting the appropriate reliability coefficient for correction. Schmidt and Hunter (2015a) then explain the correction of attenuation due to different forms of range restriction (i.e., the sample is more homogeneous in an included construct compared to the general population) or range enhancement (i.e., the sample is more heterogeneous in an included construct compared to the general population). Next, the authors explain all other artifacts, in turn, describing both the underlying mathematics and the relevance of the respective artifact for conducting a meta-analysis. However, note that not all artifacts mentioned are equally relevant in all meta-analyses, and it is not uncommon that a meta-analysis only corrects for one or two sources of bias. For example, the dichotomization of continuous variables is not an issue when all relevant primary studies report Pearson correlation coefficients.

We recommend this source to develop an understanding of the various artifacts that may attenuate meta-analytic effect sizes. Most current meta-analytic software packages contain statistical techniques to account for study artifacts. Based on the work by Schmidt and Hunter (2015a), meta-analysts can make an informed decision about which of these techniques to use.

Further reading. Most primary studies do not report sufficiently detailed information to correct all artifacts for each correlation individually. In a different chapter of their seminal work on meta-analysis, Schmidt and Hunter (2015b) describe artifact distribution meta-analysis, which may be used to correct for artifacts when not all relevant information is provided. Another recently published article by Wiernik and Dahlke (2020) discusses methods to account for different psychometric artifacts, including measurement error and selection bias in different types of meta-analyses (e.g., meta-analyses on correlations and meta-analyses on group differences). The authors provide sample R code to implement these methods.

How to Conduct the Statistical Analysis in R?

Source Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2017). A review of meta-analysis packages in R. *Journal of Educational and Behavioral Statistics*, 42, 206–242. <https://doi.org/10.3102/1076998616674315>

After extraction, transformation, and correction of the effect sizes, the actual meta-analytic aggregation can begin. A frequently used statistical software tool for this is R (R Core Team, 2020). R is an environment for statistical computing, including meta-analyses. Unlike other statistical software, R is free of charge, and anyone may contribute new statistical features in the form of “packages”. These packages contain a

collection of statistical functions that are usually well-tested and well-documented. In their manuscript, Polanin et al. (2017) provide a comprehensive review of 63 packages specific for conducting meta-analyses. First, Polanin et al. (2017) provide a relatively general narrative overview of meta-analytic techniques used in different popular R packages. The authors go on to describe the search for and documentation of relevant R packages across different sources. Table 1 in Polanin et al. (2017) summarizes the functionality of the 63 identified packages for meta-analysis and may serve as a starting point for researchers addressing a specific meta-analytic question. An alternative starting point may be the subsequent description of the packages, which is organized according to their principal functionality: general meta-analytic packages, genome meta-analytic packages, multivariate meta-analytic packages, diagnostic test accuracy meta-analytic packages, network meta-analysis packages, Bayesian meta-analysis packages, assessment of bias packages, packages with (other) specific functionality, and GUI applications. Following this overview, Polanin et al. (2017) provide two detailed tutorials for conducting meta-analyses in R, including all steps from package installation to meta-analytic effect size estimation and moderator analyses. These tutorials describe a meta-analysis with independent effect sizes in the R package *metafor* (Viechtbauer, 2010) and a meta-analysis with dependent effect sizes in the R package *robumeta* (Fisher & Tipton, 2015). The final discussion of future avenues for meta-analysis packages in R is of little relevance for beginners. Although Polanin et al. (2017) provide a comprehensive overview, the number of packages for meta-analysis is growing steadily, and developers update existing packages increasing their functionality. Thus, we advise researchers to consult the Comprehensive R Archive Network (CRAN; <http://www.cran.r-project.org>) for recent updates in meta-analysis packages for R. Please also note that a function accounting for the dependency of effect sizes from one study has recently become available in the R package *metafor*. Beginners should not be intimidated by the high number of available packages, many of which serve specialized purposes. When beginners conduct their first simple meta-analysis, mastery of a single package (oftentimes *metafor* or *robumeta*) is all that is needed.

Further reading. Most R packages for meta-analysis focus on statistical computation after the relevant effect sizes have been extracted from the literature. However, screening articles for their relevance is perhaps the most time-consuming stage of a meta-analysis. Recently, the R package *revtools* was introduced that supports data import, deduplication, and article screening for meta-analyses and systematic reviews (Westgate, 2019). Whereas previous packages that support article screening required a good knowledge of R (e.g., *METAGEAR*; Lajeunesse, 2016), *revtools* provides an easy-to-use interface for beginners.

How to Analyze Hierarchically Nested or Multivariate Effect Size Data?

Source Cheung (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 29, 387–396. <https://doi.org/10.1007/s11065-019-09415-6>

“Classical” meta-analytic approaches require that all effect sizes are mutually statistically independent. This is oftentimes not the case because primary studies may report multiple effect sizes that are relevant to a meta-analysis. This can, for example, occur when researchers use multiple indicators for a treatment outcome (e.g., psychological and physical improvements). Similarly, a study may report multiple comparisons between groups (e.g., between the control group and two different experimental groups) or a correlation coefficient for different samples (e.g., students and older adults). Cheung (2019) refers to these cases as “multivariate effect sizes”. Additionally, effect sizes may be non-independent because of other higher-level units in which the effect sizes are clustered. For example, effect sizes from studies conducted by the same research group may be more similar to each other compared to effect sizes from studies conducted by other research groups. Cheung (2019) refers to this case as “nested effect sizes”. In all of these different cases of non-independent effect sizes in meta-analyses, the statistical non-independence leads to underestimated standard errors, too narrow confidence intervals, and, ultimately, incorrect statistical inferences. The R packages *metafor* and *robumeta* offer easy-to-use functions to handle the non-independence of effect sizes and to avoid the resulting problems.

However, in some cases, researchers want to explicitly investigate how the multivariate or nested nature of the effect sizes affects the results. Cheung (2019) introduces two methodological approaches for this purpose: a multivariate meta-analysis and a three-level meta-analysis. Multivariate meta-analyses may handle multiple outcomes (similar to multivariate analyses of variance). Three-level meta-analyses take into account the variance explained by Level-2 units (e.g., study) and Level-3 units (e.g., research group). Cheung (2019) also compares the assumptions of multivariate meta-analysis and three-level meta-analyses, providing advice on when to use which meta-analytic technique. In the following most practical section of the article, Cheung (2019) briefly introduces the statistical software required to conduct multivariate meta-analyses and three-level meta-analyses (i.e., the *metaSEM* package; Cheung (2015a, 2015b) in R (R Core Team, 2020) and *Mplus* (Muthén & Muthén, 1998–2017), and illustrates advantages of the two methods over simpler meta-analytic procedure using two examples. In the last section of his article, Cheung (2019) summarizes his main points and points to future directions. In sum, Cheung’s article is a comprehensible introduction to meta-analysis with non-independent effect

sizes directed at applied researchers in psychology and beyond.

Further reading. Cheung (2019) only briefly mentions a third viable method to account for the non-independence of effect sizes: robust variance estimation (Hedges et al., 2010; Tanner-Smith et al., 2016; Tanner-Smith & Tipton, 2014). This method accounts for the non-independence of effect sizes by computing adjusted (robust) standard errors, resulting in unbiased parameter estimates (Moeyaert et al., 2017). We recommend the article by Tanner-Smith et al. (2016), which provides a tutorial for meta-analytic robust variance estimation in R (R Core Team, 2020). Practically-oriented meta-analysts may safely skip some mathematical details in the introduction section of the article. A comparison of robust variance estimation and three-level modeling is provided by López-López et al. (2017).

How Can the Heterogeneity of Effect Sizes in a Meta-Analysis be Quantified?

Source Borenstein et al. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8, 5–18. <https://doi.org/10.1002/jrsm.1230>

A central aim of meta-analyses is to reveal to which extent effect sizes from the aggregated primary studies vary around the meta-analytic mean effect size. This variance is referred to as heterogeneity. In their article, Borenstein et al. (2017) explain conceptually which measures are suitable for quantifying heterogeneity in meta-analyses and which are not. Although their manuscript primarily focuses on I^2 , Borenstein et al. (2017) also discuss alternative measures and derive clear recommendations for assessing heterogeneity. According to Borenstein et al. (2017), the main problem with I^2 is that it is usually interpreted as an absolute measure of heterogeneity. Conceptually, however, I^2 indicates what proportion of variance in the observed effects is due to variability in the true effects (i.e., variability without sampling error). This proportion—by definition—can not indicate how much the effects vary (in terms of absolute value). Borenstein et al. (2017) thoroughly explain this point conceptually and illustratively. Moreover, Borenstein et al. (2017) describe prediction intervals as a more suitable alternative to I^2 . They define the 95% prediction interval (PI) as the range of $\pm$ two standard deviations of the true effects around their common mean. If the effects are normally distributed, a total of 95% of effect sizes of all populations will fall in this range. Why are prediction intervals thus appealing? If a meta-analyst is asked to predict the effect size (e.g., the efficacy of treatment) for any population (that is similar to the population included in the meta-analysis), they could predict that the effect size would fall in this range and would be correct with this prediction 95% of the time. Prediction intervals are also important if

one uses meta-analytic results for power calculations for a new study. The expected true effect in this new study can be any of the values in the prediction interval and is not necessarily the estimated mean effect size of the meta-analysis (Int'Hout et al., 2016). The article additionally explains prediction intervals for ratios, prevalence, and correlations. Additionally, the authors provide a spreadsheet for the computation of prediction intervals. In the last sections of their manuscript, Borenstein et al. (2017) explain when I^2 might be informative (i.e., to indicate to which extent measurement error explains heterogeneity) and illustrate how prediction intervals provide relevant information beyond I^2 . The appendix of Borenstein et al. (2017) provides an additional illustration for computing different measures associated with heterogeneity and their interpretation. Overall, the article provides an accessible conceptual introduction into assessing heterogeneity in meta-analyses, including step-by-step guidance to computing and interpreting prediction intervals.

Further reading. We additionally recommend the seminal article by Higgins and Thompson (2002), in which the authors discuss the shortcomings of previously used measures of heterogeneity in meta-analyses. However, keep in mind the possible misinterpretations of I^2 highlighted by Borenstein et al. (2017) when reading Higgins and Thompson (2002). Researchers with interest in Bayesian meta-analysis should read the article by Turner et al. (2015) that introduces methods for the quantification of heterogeneity in Bayesian meta-analysis. These methods allow the inclusion of external evidence for the expected magnitude of heterogeneity in a meta-analysis, which ultimately leads to more precise estimates of between-study heterogeneity.

How to Detect and Quantify Publication Bias in Meta-Analyses?

Source 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, 115–144. <https://doi.org/10.1177/2515245919847196>

Meta-analytical results and conclusions can be biased when the included published studies are systematically unrepresentative of all studies conducted on a research topic. Meta-analysts describe this phenomenon with the term “publication bias”. This bias may, for example, occur when significant results are easier to publish than non-significant results (*file drawer problem*). Numerous researchers have developed statistical techniques to detect and quantify publication bias and to ultimately correct meta-analytical results for it. Carter et al. (2019) compared seven different detection and correction techniques for publication bias (e.g., trim-and-fill, p -curve, p -uniform, PET, PEESE, and PET-PEESE) using 1000

meta-analyses based on simulated data. The data sets included varying conditions regarding the degree of heterogeneity, the degree of publication bias, and the use of questionable research practices such as the optional exclusion of outliers or optional stopping. The authors showed that none of the seven tested methods performed acceptably under all conditions. They consequently recommend that meta-analysts in psychology should conduct sensitivity analysis or method performance checks, that is, applying multiple publication bias methods and consider the conditions under which these methods tend to fail.

Further reading. A practical guide on how to conduct different types of publication bias analyses in R is provided online by Viechtbauer (2020). In this online resource, he provides the R code to reproduce the worked examples and analyses presented in the book *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* by Rothstein et al. (2005). For those interested in performing a meta-analysis in R, we recommend also browsing the other valuable online resources of Viechtbauer (2020).

How Should a Meta-Analysis be Reported?

Source American Psychological Association (APA). (2020). *Quantitative Meta-Analysis Article Reporting Standards - Information recommended for inclusion in manuscripts reporting quantitative meta-analyses*. <https://apastyle.apa.org/jars/quant-table-9.pdf>

Researchers who usually publish primary studies may experience writing up a meta-analysis as a difficult task. In their Journal Article Reporting Standards, the American Psychological Association (APA, 2020) provides detailed recommendations regarding the information to be included in a manuscript reporting meta-analytic findings. These recommendations pertain to all parts of the manuscript, from the title and title page to the discussion section. The APA (2020) has provided their recommendations in a concise table so that they are easily digestible. In principle, all of the aspects mentioned should be covered in any manuscript reporting findings from a quantitative meta-analysis. However, due to space constraints, some of the information required may have to be moved to the supplementary material (e.g., a table displaying the characteristics of each included study). Similarly, some aspects required for the abstract could go beyond the word limit of some journals (e.g., results for moderator analyses) and might, therefore, only be briefly touched. In sum, we recommend that meta-analysts carefully study and follow the recent APA (2020) Reporting Standards. This will greatly facilitate crafting a manuscript and reporting findings from a meta-analysis.

Further reading. Cooper (2019) provides additional advice on the APA Meta-Analysis Reporting Standards, including

examples from APA journals and instructions for implementing APA standards in one's writing. Some journals in Psychology and neighboring scientific disciplines require authors to provide Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) statement along with their manuscript when submitting a meta-analysis. The PRISMA statement contains a checklist and a flow diagram. In the checklist, meta-analysts indicate where in their submitted manuscript they have reported essential information requested by PRISMA. The PRISMA flow diagram is a figure used to provide an overview of the articles' identification, screening, and inclusion/exclusion process. There is also an excellent overview of guides and standards for reporting meta-analyses provided by the Countway Library of Medicine (2021).

How to Conduct a Meta-Analysis Considering Open Science Practices?

Source Moreau, D., & Gamble, B. (2020). Conducting a meta-analysis in the age of open science: Tools, tips, and practical recommendations. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000351>

Researchers have to make many decisions when conducting a meta-analysis that can greatly influence overall estimates of effect sizes, their statistical significance, or subsequent subgroup analyses. Open science practices (i.e., pre-registration, open material, open data) serve at least three major purposes. First, preregistration helps to safeguard oneself against questionable research practices and provides a record of the original plan of a study. Second, open material (e.g., R scripts for data preparation and analysis) can promote the reproducibility of the meta-analytic results. Third, open data (e.g., metadata about each included article, descriptive statistics, notes, and additional information for each study) can provide the basis for further reviews and meta-analyses and increases the transparency about the meta-analytical findings, and again facilitates reproducibility. The tutorial by Moreau and Gamble (2020) serves as an excellent starting point for meta-analysts aiming to conduct transparent, robust, and reproducible research by embracing open science values. The tutorial starts with an introduction to open science more generally and then focuses on specific recommendations regarding meta-analyses (e.g., a comparison of two platforms (Open Science Framework and PROSPERO) for preregistrations taking into account aspects important for meta-analysts. Especially valuable in this tutorial is the excessive online materials, which contain nine useful templates. Templates 1 and 2 are thought to support meta-analysts with the preparation of a preregistration following the PRISMA protocol. Template 3 includes a list of R packages that can be used for data wrangling and data visualization in the context of meta-analyses.

Template 4 provides a search syntax template to enhance the reproducibility of the standardized literature search that can be adopted for different databases. Templates 5 and 6 are email templates that can guide meta-analysts by sending out a call for unpublished data during the initial stage of the search or by requesting information from corresponding authors of relevant articles that lack information or need clarification. Template 7 was designed to help meta-analysts sharing important information on the literature search results (e.g., number of entries, all references, and abstracts). Template 8 can help to openly share the actual meta-analytical database, including metadata about each included article, such as descriptive statistics, sample characteristics, and notes for each study. Last, Template 9 guides meta-analysts that deviated from their initial preregistration—a very common phenomenon. This template includes examples of how to document and justify such deviations using examples from applied psychology.

Further reading. We additionally recommend the article on the reproducibility of meta-analyses by Lakens et al. (2016). The authors provide six practical recommendations to increase the credibility of meta-analytic conclusions and to allow updates of meta-analyses after several years.

Limitations

This annotated reading list has noteworthy limitations. First, and most importantly, the inclusion of resources in this list is subjective. We have attempted to provide a broad overview of meta-analytic resources and topics. Yet, our inherent views will have biased the reading list composition, and other researchers in the field may have chosen different foci. Second, we could not address all current developments in meta-analysis in this article. The interested reader will find various recent methodological developments, such as meta-analytic structural equation modeling (Cheung, 2015a) or Bayesian model-averaged meta-analysis (Gronau et al., 2020), for which the literature presented here provides a profound basis. Third, in this article, we focused on literature providing guidance for conducting meta-analyses in the R statistical environment. We chose this focus because R is a free software for which a large and active community continuously provides new packages that incorporate the newest developments in meta-analytic techniques.

Conclusions

In this manuscript, we aimed to provide an accessible overview of relevant literature for researchers that aim to conduct a meta-analysis in psychology. This annotated reading list covered classical readings and recent publications on state-of-the-art meta-analytic techniques. After consulting the literature

presented here, researchers should understand the basic concepts of conducting and reporting a meta-analysis in psychology. Yet, meta-analysis is a quickly growing field, and new techniques and software become available in rapid intervals. Thus, based on the principal insights into meta-analysis gained through the literature presented here, we recommend that researchers interested in meta-analysis frequently update their knowledge on current issues and new developments regarding meta-analytical methods, software, or reporting standards. A valuable resource for this purpose might be Evidence Synthesis Hackathon (<https://www.eshackathon.org>).

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study. Consequently, also no statement of informed consent was necessary. The authors have no conflicts of interest to declare. The present study does not use any datasets and therefore was exempt from approval by the Ethics Committees of the author's institutions (Ruhr University Bochum, Germany and University of Trier, Germany), in accordance with national law.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aloe, A. M. (2015). Inaccuracy of regression results in replacing bivariate correlations: Inaccuracy of regression results. *Research Synthesis Methods*, 6, 21–27. <https://doi.org/10.1002/jrsm.1126>.
- American Psychological Association (APA). (2020). *Quantitative Meta-Analysis Article Reporting Standards—Information recommended for inclusion in manuscripts reporting quantitative meta-analyses.* <https://apastyle.apa.org/jars/quant-table-9.pdf>. Accessed 20 June 2021.
- Baumeister, R. F. (2013). Writing a literature review. In M. J. Prinstein (Ed.), *The portable mentor* (pp. 119–132). Springer. https://doi.org/10.1007/978-1-4614-3994-3_8.
- Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, 1, 311–320. <https://doi.org/10.1037/1089-2680.1.3.311>.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009a). How a meta-analysis works. In: *Introduction to meta-analysis* (pp. 3–7). John Wiley & Sons. <https://doi.org/10.1002/9780470743386.ch1>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009b). *Introduction to Meta-analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470743386>.

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009c). Preface. In: *Introduction to meta-analysis* (pp. xxi–xxvi). John Wiley & Sons.
- Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8, 5–18. <https://doi.org/10.1002/jrsm.1230>.
- Calhoun, C. D. (2013, October). *Finding what you need: Tips for using PsycINFO effectively*. <https://www.apa.org/science/about/psa/2013/10/using-psycinfo>. Accessed 20 June 2021.
- 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, 115–144. <https://doi.org/10.1177/2515245919847196>.
- Chan, M. E., & Arvey, R. D. (2012). Meta-analysis and the development of knowledge. *Perspectives on Psychological Science*, 7, 79–92. <https://doi.org/10.1177/1745691611429355>.
- Cheung, M. W.-L. (2015a). *Meta-analysis: A structural equation modeling approach*. Wiley.
- Cheung, M. W.-L. (2015b). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, 5, 5. <https://doi.org/10.3389/fpsyg.2014.01521>.
- Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 29, 387–396. <https://doi.org/10.1007/s11065-019-09415-6>.
- Cheung, M. W.-L., & Vijayakumar, R. (2016). A guide to conducting a meta-analysis. *Neuropsychology Review*, 26, 121–128. <https://doi.org/10.1007/s11065-016-9319-z>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.
- Cooper, H. M. (2019). *Reporting quantitative research in psychology: How to meet apa style journal article reporting standards* (second edition, revised). American Psychological Association.
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.
- Countway Library of Medicine. (2021). *Systematic reviews and meta analysis: Guides and standards*. <https://guides.library.harvard.edu/meta-analysis/guides>. Accessed 20 June 2021.
- Del Re, A. C. (2013). *compute.es: Compute effect sizes* (R package version 0.2-2) [computer software]. <https://cran.r-project.org/package=compute.es>. Accessed 20 June 2021.
- Doi, S. A. R., Barendregt, J. J., Khan, S., Thalib, L., & Williams, G. M. (2015). Simulation comparison of the quality effects and random effects methods of meta-analysis. *Epidemiology*, 26, e42–e44. <https://doi.org/10.1097/EDE.0000000000000289>.
- Field, A. P. (2005). Meta-analysis. In J. Miles & P. Gilbert (Eds.), *A handbook of research methods for clinical and health psychology* (pp. 295–308). Oxford University Press. <https://doi.org/10.1093/med:psych/9780198527565.001.0001>.
- Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. *ArXiv:1503.02220 [Stat]*. <http://arxiv.org/abs/1503.02220>. Accessed 20 June 2021.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18. <https://doi.org/10.1037/a0024338>.
- Gronau, Q. F., Heck, D. W., Berkhout, S. W., Haaf, J. M., & Wagenmakers, E.-J. (2020). *A primer on Bayesian model-averaged meta-analysis*. PsyArXiv. <https://doi.org/10.31234/osf.io/97qup>.
- Harari, M. B., Parola, H. R., Hartwell, C. J., & Riegelman, A. (2020). Literature searches in systematic reviews and meta-analyses: A review, evaluation, and recommendations. *Journal of Vocational Behavior*, 118, 103377. <https://doi.org/10.1016/j.jvb.2020.103377>.
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426–445. <https://doi.org/10.1037/1082-989X.9.4.426>.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1, 39–65. <https://doi.org/10.1002/jrsm.5>.
- Higgins, J. P. T., & Cochrane Collaboration (Eds.). (2020). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Wiley-Blackwell.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558. <https://doi.org/10.1002/sim.1186>.
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8, 275–292. <https://doi.org/10.1111/1468-2389.00156>.
- IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6, e010247. <https://doi.org/10.1136/bmjopen-2015-010247>.
- Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis: Power analysis for meta-analysis. *Research Synthesis Methods*, 8, 290–302. <https://doi.org/10.1002/jrsm.1240>.
- Lajeunesse, M. J. (2016). Facilitating systematic reviews, data extraction and meta-analysis with the METAGEAR package for R. *Methods in Ecology and Evolution*, 7, 323–330. <https://doi.org/10.1111/2041-210X.12472>.
- Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, 4, 24. <https://doi.org/10.1186/s40359-016-0126-3>.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.
- López-López, J. A., Van den Noortgate, W., Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2017). Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: A Monte Carlo simulation. *Research Synthesis Methods*, 8, 435–450. <https://doi.org/10.1002/jrsm.1245>.
- Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, 20, 559–572. <https://doi.org/10.1080/13645579.2016.1252189>.
- 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, e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- Moreau, D., & Gamble, B. (2020). Conducting a meta-analysis in the age of open science: Tools, tips, and practical recommendations. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000351>.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Othman, R., & Sahlawaty Halim, N. (2004). Retrieval features for online databases: Common, unique, and expected. *Online Information Review*, 28, 200–210. <https://doi.org/10.1108/14684520410543643>.
- Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2017). A review of meta-analysis packages in R. *Journal of Educational and Behavioral Statistics*, 42, 206–242. <https://doi.org/10.3102/1076998616674315>.
- Pustejovsky, J. E. (2014). Converting from d to r to z when the design uses extreme groups, dichotomization, or experimental control. *Psychological Methods*, 19, 92–112. <https://doi.org/10.1037/a0033788>.

- R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. <https://www.r-project.org/>. Accessed 20 June 2021.
- Rothstein, H., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley.
- Schmidt, F. L., & Hunter, J. E. (2015a). Meta-analysis of correlations corrected individually for artifacts. In: *Methods of meta-analysis: Correcting error and bias in research findings* (pp. 87–164). SAGE publications. <https://doi.org/10.4135/9781483398105>
- Schmidt, F. L., & Hunter, J. E. (2015b). Meta-analysis of correlations using artifact distributions. In: *Methods of meta-analysis: Correcting error and bias in research findings* (pp. 165–212). SAGE Publications. <https://doi.org/10.4135/9781483398105>
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70, 747–770. <https://doi.org/10.1146/annurev-psych-010418-102803>.
- Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, 5, 13–30. <https://doi.org/10.1002/jrsm.1091>.
- Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2, 85–112. <https://doi.org/10.1007/s40865-016-0026-5>.
- Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., & Higgins, J. P. T. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, 34, 984–998. <https://doi.org/10.1002/sim.6381>.
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need?: A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35, 215–247. <https://doi.org/10.3102/1076998609346961>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. <https://doi.org/10.1103/PhysRevB.91.121108>.
- Viechtbauer, W. (2020). R code corresponding to the book publication Bias in Meta-analysis by Rothstein et al. (2005). https://wviechtb.github.io/meta_analysis_books/rothstein2005.html. Accessed 20 June 2021.
- Westgate, M. J. (2019). Revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods*, 10, 606–614. <https://doi.org/10.1002/jrsm.1374>.
- Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, 3, 94–123. <https://doi.org/10.1177/2515245919885611>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.