ORIGINAL MANUSCRIPT

What do meta-analysts need in primary studies? Guidelines and the SEMI checklist for facilitating cumulative knowledge

Belén Fernández-Castilla[1] · Sameh Said-Metwaly[2,3,4] · Rodrigo S. Kreitchmann[1] · Wim Van Den Noortgate[2,3]

Abstract

Meta-analysis is often recognized as the highest level of evidence due to its notable advantages. Therefore, ensuring the precision of its findings is of utmost importance. Insufficient reporting in primary studies poses challenges for meta-analysts, hindering study identification, effect size estimation, and meta-regression analyses. This manuscript provides concise guidelines for the comprehensive reporting of qualitative and quantitative aspects in primary studies. Adhering to these guidelines may help researchers enhance the quality of their studies and increase their eligibility for inclusion in future research syntheses, thereby enhancing research synthesis quality. Recommendations include incorporating relevant terms in titles and abstracts to facilitate study retrieval and reporting sufficient data for effect size calculation. Additionally, a new checklist is introduced to help applied researchers thoroughly report various aspects of their studies.

Keywords Meta-analysis · Cumulative knowledge · Checklist

Meta-analysis is a statistical technique that emerged in response to the need to combine results from studies addressing similar research questions to draw a general conclusion about the state-of-the-art of a given research topic (Glass, 1976). This methodology began to be implemented in the 1980s when it was uncommon for authors to make the datasets utilized in their studies freely available. The difficulty in accessing raw data led to the need to use the results reported in each study to obtain a quantitative measure of the strength of the effect of interest, namely the effect size measure (Glass et al., 1981; Ray & Shadish, 1996).

Belén Fernández-Castilla and Sameh Said-Metwaly equally contributed to the work.

✉ Belén Fernández-Castilla
bfcastilla@psi.uned.es

[1] Faculty of Psychology, Universidad Nacional de Educación a Distancia, Juan del Rosal 10, 28040 Madrid, Spain

[2] Faculty of Psychology and Educational Sciences, KU, Leuven, Belgium

[3] Imec-Itec, KU, Leuven, Belgium

[4] Faculty of Education, Damanhour University, Damanhour, Egypt

The introduction of meta-analysis as a research synthesis technique has led to several potential advantages. Meta-analyses rely on replicable, transparent, and inclusive methodology to identify relevant studies (encompassing not only peer-reviewed results but also pertinent gray literature; Rytwinski et al., 2021). By accumulating data from multiple studies, a meta-analysis allows for more accurate estimation of the overall effect size, maximizing the statistical power and generalizability of the effect size, assessing heterogeneity across studies and explaining it through moderator variables, answering questions not researched in individual studies, developing hypotheses for future consideration, and permitting a regular update of results using newly available data (Deeks et al., 2008; Egger & Smith, 1997; Haidich, 2010; Walker et al., 2008). For these reasons, meta-analysis is frequently considered the highest rank in the hierarchy of evidence (Cooper et al., 2019), implying greater trust in its results than in those of primary studies. This underscores the importance of ensuring that the results of meta-analyses are as reliable and valid as possible.

While meta-analysis is a valuable methodology, it poses a significant challenge due to the considerable time it demands. The process involves searching, screening, and extracting data from all relevant studies, calculating effect sizes and corresponding sampling variances, and carrying out statistical analyses (i.e., syntheses of effect sizes and

meta-regression analyses, Cooper et al., 2019). Each step is time-consuming, and complications arise when relevant primary study information is not (clearly) reported. For instance, if the variables of interest are expressed in ambiguous terms in the title or abstract, meta-analysts might have difficulties retrieving that study. Also, insufficient reporting of key study characteristics (e.g., related to sample, design, or setting) hinders meta-analysts' ability to extract relevant information and incorporate it into meta-regression analyses. Primary researchers might also fail to report quantitative information essential for meta-analysts to calculate effect sizes. Hence, proper reporting of various aspects of primary studies can facilitate more efficient work for meta-analysts, leading to thorough and rigorous research syntheses. Since primary researchers may not always be aware of the information required by a meta-analyst for integrating their study into a research synthesis, the goal of this manuscript is to offer concise instructions on reporting both qualitative and quantitative aspects of primary research. This will enable primary researchers to improve the eligibility of their studies for inclusion in future research syntheses, ultimately resulting in heightened visibility and impact within the academic community and society.

Numerous guidelines for conducting and reporting quantitative research are available and endorsed (e.g., Appelbaum et al., 2018). Adhering to these guidelines can enhance the overall quality of a study. However, it must be noted that improved quality does not necessarily guarantee eligibility for inclusion in a meta-analysis. Meta-analysis criteria often involve additional considerations beyond individual study quality, emphasizing factors such as data relevance and sufficiency. Therefore, meeting guidelines is a valuable step, but researchers should be mindful of the distinct requirements for meta-analytic eligibility. In addressing this issue, Chow et al. (2023) introduced guidelines, with a strong focus on open science. While our study incorporates several of their guidelines, we also introduce supplementary ones not covered by Chow et al. (2023). For instance, we emphasize the role of thorough reporting in aiding various steps of a meta-analysis, including study searching and screening, as well as effect size estimation. Additionally, while we also acknowledge the value of open data, there are instances where sharing data may not always be feasible or may not necessarily enable meta-analysts to retrieve the information needed for research synthesis. Therefore, unlike Chow et al. (2023), we place special emphasis on reporting readily available relevant statistics to streamline the meta-analyst's workflow and enhance clarity for all report users.

Through the remainder of this document, we outline the stages associated with conducting a meta-analysis, focusing particularly on those stages directly influenced by the quality of reporting in primary studies. At each stage, we highlight the essential components that need to be incorporated into primary studies to enable future meta-analyses. Furthermore, we discuss the significant role that open science practices play in incorporating a specific study into a research synthesis. Ultimately, we present the Study Eligibility for Meta-Analysis Inclusion (SEMI) checklist, offering concise and clear reporting guidelines for applied researchers to enhance the potential inclusion of their studies in a meta-analysis.

Searching and screening the literature

In general, a meta-analysis commences with a systematic literature search. Researchers select a set of keywords to search electronic databases for relevant studies. The omission of a crucial keyword may result in overlooking valuable studies in the meta-analysis (Alexander, 2020). The selection of these keywords thus holds substantial significance as it directly influences the number of studies retrieved and may induce bias in the meta-analytic dataset.

In an ideal scenario, meta-analysts would conduct an extensive search for the keywords throughout the full text of research papers. However, if the scope of the meta-analytic investigation or the keyword list is broad, an overwhelming quantity of potentially relevant studies may surface, and many of them may prove irrelevant. To streamline the search process, a commonly employed strategy involves restricting the keyword search to the study title and abstract, assuming that authors normally indicate the most pertinent information within these sections. In this regard, our first recommendation is that authors always clearly mention the most relevant variables under investigation and study characteristics in the study title and abstract so that their study can be easily located during the search phase (aligning with APA reporting standards, see Appelbaum et al., 2018).

Another approach to identifying pertinent studies involves a backward search, wherein references cited within studies are examined. Primary studies, which effectively provide a general overview of the most important literature on the topic and extend beyond the studies published or indexed in databases, serve as a valuable source of studies for meta-analysis. They contribute to the discovery of additional relevant studies that may not have been initially located through conventional database searches, preventing potential oversights in the search process.

Once the researcher has compiled a list of all potential studies, a subsequent step involves the initial screening phase. Based on predetermined inclusion criteria, the researcher (or a group of researchers) assesses the relevance of studies based on their titles and abstracts, excluding those that do not meet the criteria. To expedite the screening process, the title should be as informative as possible and an abstract should capture essential details about a study, offering an accurate record of its conduct and results within

the space constraints of a journal (Appelbaum et al., 2018; Polanin et al., 2019). In cases where the title and abstract do not conclusively establish a study's relevance, a meta-analyst is compelled to delve into the full text. Thus, a clear presentation of research objectives or questions and research outcomes within the study is crucial for a swift determination of its relevance.

Coding the literature

Upon selecting the studies for inclusion in the research synthesis, the next step entails extracting the pertinent qualitative and quantitative information from each study. This information serves three primary purposes: (1) qualitatively summarizing the characteristics of the included studies, (2) quantitatively calculating the desired effect sizes, and (3) conducting moderator analyses, wherein study characteristics (referred to as moderator variables) are employed in a meta-regression model to examine their relationship with the observed effect sizes.

One common challenge in this phase is the incomplete reporting of study characteristics and/or insufficient data within studies to compute the effect size (Lee & Beretvas, 2022; Pigott, 2019; Tipton et al., 2019), which can lead to study exclusion from the meta-analysis (or from moderator analyses) and consequently impact statistical power. Hence, we urge researchers to follow the next guidelines and report study characteristics and outcomes in sufficient detail so that these aspects can be easily coded and used in future research synthesis.

To identify specific study characteristics relevant to future research synthesis, particularly for moderator analyses, the PICO framework (McGowan et al., 2016) can be employed. In this framework, P refers to participant characteristics (e.g., number, age, gender, or socioeconomic status), I refers to intervention or exposure details (e.g., experimental condition, modality, duration, or medication type), C refers to comparator characteristics (e.g., control condition such as a traditional treatment or waitlist), and O refers to outcome characteristics (e.g., a comprehensive description of dependent variables). Alternative frameworks, such as SPICE (setting, perspective, intervention, comparison, evaluation; Booth, 2006) and SPIDER (sample, phenomenon of interest, design, evaluation, research type; Cooke et al., 2012), can also be applied across various study designs.

Within all these frameworks, it is recommended to provide a comprehensive and accurate description of the sample, particularly highlighting characteristics that may impact the results. These include details such as the number of participants identifying as men and women, mean age (including standard deviation or range), region of origin, or socioeconomic status. For research synthesis

purposes, authors are urged to present this information on the final analyzed sample, specifically after dropout removal, which may vary across analyses within the same study. Additionally, citing other studies utilizing the same sample or subset thereof is vital to preventing overrepresentation and ensure the unique contributions of samples in meta-analyses. Furthermore, avoiding duplication of samples in meta-analyses is essential for maintaining statistical independence among studies, which is crucial for accurate meta-analytic estimates.

These frameworks also emphasize the necessity of appropriately describing independent and dependent variable(s). For independent variables, such as interventions or experimental conditions (as seen in the PICO or SPICE framework), primary studies should include crucial information such as the modality and intensity of the intervention/experimental condition, its duration (number of sessions and session duration), and details on any administered drugs and their quantities. In correlational studies, where the independent variable is observed rather than experimentally manipulated, it is imperative to furnish information on how the independent variable is operationalized, measured (including reliability measures calculated on the observed data), and implemented. The same level of detail is essential for the dependent variables. Including these specifics not only increases the likelihood that a study can be included in a meta-analysis but also enables the assessment of its risk of bias.

Methodological characteristics are also crucial for meta-analysts to evaluate the methodological quality of primary studies (Pigott & Polalin, 2020). These include the specific research design (e.g., experimental, quasi-experimental, cross-sectional, or longitudinal), procedural details (e.g., where, how, and when data are collected, and the randomization of participants across groups), and specifics of the data-analytic methods (e.g., significance level, statistical tests, and whether the test is two-sided or one-sided).

In terms of reporting the methodological aspects of a study, applied researchers can utilize relevant risk-of-bias assessment tools for comprehensive reporting (refer to https://www.latitudes-network.org/ for an overview of pertinent risk-of-bias assessment tools, Whiting et al., 2023). For example, the widely used Risk of Bias Tool 2 (RoB2; Sterne et al., 2019) for assessing randomized controlled trials includes items such as evaluating the randomization process and assessing bias due to deviations from intended interventions. Authors of primary studies should accurately describe participant assignment and provide specific details on blinding and potential deviations from intended therapy. Systematically reviewing various items from diverse risk-of-bias assessment tools, available for different research designs, significantly assists applied researchers in providing necessary information for others to assess the quality of their studies.

As previously mentioned, proper reporting of numerical results is essential for calculating commonly used effect sizes. The following section provides a brief overview of the statistical outcomes required for effect size calculations.

Calculating and combining study outcomes

The next step in a meta-analysis involves calculating an index that summarizes the strength of the effect of interest targeted for meta-analysis. Commonly known as an effect size measure, it is defined as "*a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest*" (Kelley & Preacher, 2012, pp. 140). However, we do not recommend simply reporting effect sizes that address the research questions of the primary study. This is because the meta-analyst may be interested in an effect size associated with a different set of variables within that study. For example, consider a study with the aim of investigating the effectiveness of an intervention on two dependent variables: well-being and anxiety symptoms. The authors may report two Cohen's *d* values summarizing the intervention's effectiveness, successfully addressing the intended effect size in that study. However, the meta-analyst might be interested in the correlation between well-being and anxiety symptoms. Since this correlation might not be the primary focus for the primary authors, it might be overlooked in their reporting, consequently leading to the exclusion of the study from meta-analysis. This exclusion can be avoided if authors are contacted to share the correlation value or if they make the dataset publicly available on an online repository, enabling meta-analysts to calculate any desired effect size related to the studied variables. Another reason why merely reporting effect sizes may not be sufficient for a study to qualify for inclusion in a meta-analysis is that for certain types of effect sizes, different formulas exist (e.g., Cohen's *d* in correlated samples, more information is given in subsequent sections) that might represent different, incomparable parameters (Lakens, 2013). If the authors do not explicitly specify the formula they employed, the meta-analyst in question will be unable to determine whether the reported effect size is appropriate for the research synthesis.

As a result, primary studies should not only report the primary effect size relevant to their specific research question but also provide the necessary numerical information to facilitate its calculation, including its precision (i.e., sampling variance). Since primary researchers might not know which numerical information future meta-analysts will need for their studies, a significant section of this manuscript outlines guidelines regarding the specific

quantitative data that should be reported. This aims to enable future meta-analysts to calculate their desired effect size, thereby facilitating the inclusion of the primary study in research synthesis.

The following sections are organized as follows: Firstly, we discuss the role of open science in research synthesis and associated barriers. Next, we attempt to unpack the information that primary investigators should provide in their papers (either in the main text or in supplementary material) to increase the likelihood that their study will be eligible for research synthesis. Although information on the calculation, reporting, and interpretation of effect sizes can be found elsewhere (e.g., Borenstein, et al., 2021; Cooper et al., 2019; Cumming, 2012; Durlak, 2009; Grissom & Kim, 2005 Lakens, 2013; Olejnik & Algina, 2000; Pek & Flora, 2018; Schmidt & Hunter, 2014; Trusty et al., 2004), in Table 1 we provide a summary of the formulas for calculating popular effect sizes to support the information stated below.

Open science

While comprehensive reporting is crucial for study eligibility in research synthesis as outlined in the following sections, the significance of this reporting may diminish if raw datasets are consistently accessible. If raw datasets are publicly available, meta-analysts could calculate any effect size of interest, whether the one reported in the study or any other beyond the primary study goal. Additionally, with raw data available in all studies, individual participant data meta-analyses (Riley et al., 2010) could be systematically performed. Hence, giving access to the datasets would undoubtedly assist meta-analysts in retrieving important data to conduct a research synthesis, namely the effect sizes and relevant information for the moderator analyses.

Despite the increasing number of journals and granting agencies mandating the sharing of collected data, the actual practice of data sharing remains relatively infrequent. Obstacles to data sharing extend beyond technical challenges. Issues such as the absence of recognition incentives for sharing research data, the absence of standardized formats for data and metadata (that offer the details necessary for other researchers to comprehend the data), privacy concerns, fear of misuse, and limited time and resources all pose potential hindrances to effective data sharing (Krumholz, 2012).

Even in cases of successful data sharing, it does not necessarily contribute to resolving reporting issues for meta-analysis. First, providing the dataset and the analytics code to reproduce the main results does not always ensure reproducibility (Hardwicke et al., 2018; Hardwicke et al., 2021; Obels et al., 2020). This is because authors may make errors in the dataset and/or code, or they may not provide

Table 1 Formulas for the calculation of commonly used effect sizes and corresponding sampling variance

Effect size measure	Study design	Statistics needed	Effect size formula	Sampling variance
Standardized mean difference	Independent groups design	$\overline{X}_1, \overline{X}_2, S_1^2, S_2^2, n_1, n_2$	$g_{ig} = \frac{\overline{X}_1 - \overline{X}_2}{S_{pooled}} \cdot \left[1 - \frac{3}{4 \cdot (n_1+n_2)-9}\right]$ where $S_{pooled} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$	$S_{g_{ig}}^2 = \left[\frac{n_1+n_2}{n_1 \cdot n_2} + \frac{g^2}{2(n_1+n_2)}\right] \cdot \left[1 - \frac{3}{4 \cdot (n_1+n_2)-9}\right]^2$
	Matched group design (1)	$\overline{X}_{pre}, \overline{X}_{post}, S_{difference}, r_{pre-post}$	$g_{rm_{(1)}} = \frac{\overline{X}_{post} - \overline{X}_{pre}}{S_{within}} \cdot \left[1 - \frac{3}{4 \cdot (n-1)-1}\right]$ where $S_{within} = \frac{S_{difference}}{\sqrt{2(1-r)}}$	$S_{g_{rm_{(1)}}}^2 = \left(\frac{1}{n} + \frac{g^2_{rm_{(1)}}}{2n}\right)(2(1-r)) \cdot \left[1 - \frac{3}{4 \cdot (n-1)-1}\right]^2$
	Matched group design (2)	$\overline{X}_{pre}, \overline{X}_{post}, S_{pre}, r_{pre-post}$	$g_{rm_{(2)}} = \frac{\overline{X}_{post} - \overline{X}_{pre}}{S_{pre}} \cdot \left[1 - \frac{3}{4 \cdot (n-1)-1}\right]$	$S_{g_{rm_{(2)}}}^2 = \left[\frac{2(1-r)}{n}\right]\left(\frac{n-1}{n-3}\right)\left[1 + \frac{ng^2_{rm_{(2)}}}{2(1-r)}\right]\left[1 - \frac{3}{4 \cdot (n-1)-1}\right]^2 - g^2_{rm_{(2)}}$
	Independent groups with pre- and post-test measures (1)	$\overline{X}_{1,pre}, \overline{X}_{2,pre}, \overline{X}_{1,post}, \overline{X}_{2,post}, S_{1,difference}, S_{2,difference}, S_{1,pre}, S_{2,pre}, r_{1,pre-post}, r_{2,pre-post}$	$g_{igpp_{(1)}} = g^E_{rm_{(1)}} - g^C_{rm_{(1)}}$	$S_{g_{igpp_{(1)}}}^2 = S^E_{g_{rm_{(1)}}} + S^C_{g_{rm_{(1)}}}$
Standardized mean difference (cont.)	Independent groups with pre- and post-test measures (2)	$\overline{X}_{1,pre}, \overline{X}_{2,pre}, \overline{X}_{1,post}, \overline{X}_{2,post}, S_{1,pre}, S_{2,pre}, r_{1,pre-post}, r_{2,pre-post}$	$g_{igpp_{(2)}} = g^E_{rm_{(2)}} - g^C_{rm_{(2)}}$	$S_{g_{igpp_{(2)}}}^2 = S^E_{g_{rm_{(2)}}} + S^C_{g_{rm_{(2)}}}$
Risk ratio	Prospective/Longitudinal studies	f_1, f_0, n_1, n_0 (see 2×2 table below)	$RR = \frac{p_1}{p_0}$ where $p_1 = \frac{f_1}{n_1} = $ and $p_0 = \frac{f_0}{n_0}$	$S^2_{Ln(RR)} = \frac{1-p_1}{p_1 \cdot n_1} + \frac{1-p_0}{p_0 \cdot n_0}$
Odds ratio	Cross-sectional	f_1, f_0, f_2, f_3 (see 2×2 table below)	$OR = \frac{f_1 \cdot f_2}{f_0 \cdot f_3}$	$S^2_{\log(OR)} = \frac{1}{f_0} + \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3}$
Psychometric effect sizes	McDonald's ω	λ_i, ε_i	$\omega = \frac{\left(\sum_{i=1}^l \lambda_i\right)^2}{\left(\sum_{i=1}^l \lambda_i\right)^2 + \sum_{i=1}^l \varepsilon_i}$	No consensus (use e.g., bootstrap)
	Validity correction for unreliability	$r_{xy}, r_{xx'}, r_{yy'}$	$\rho_{xy} = \frac{r_{xy}}{\sqrt{r_{xx'} \cdot r_{yy'}}}$	No consensus (use e.g., bootstrap)
	Validity correction for range restriction	r, r^u	$\rho^u = \frac{r}{a}$ where $a = \frac{r}{r^u}$	$S^2_{\rho^u} = \frac{S_r^2}{a^2}$

$\overline{X}_1 =$ mean of the first group; $\overline{X}_2 =$ mean of the second group; $S_1^2 =$ variance of the first group; $S_2^2 =$ variance of the second group; $\overline{X}_{pre} =$ mean of the pre-test; $\overline{X}_{post} =$ mean of the post-test; $S_{difference} =$ standard deviation of the difference scores (i.e., difference between pre-test and post-test scores); $S_{pre} =$ standard deviation of the pre-test scores; $r_{pre-post} =$ correlation between pre-test and post-test scores; $\overline{X}_{1,pre} =$ mean score of first group at pre-test; $\overline{X}_{2,pre} =$ mean score of second group at pre-test; $\overline{X}_{1,post} =$ mean score of first group at post-test; $\overline{X}_{2,post} =$ mean score of second group at post-test; $S_{1,difference} =$ standard deviation of the difference scores in the first group; $S_{2,difference} =$ standard deviation of the difference scores in the second group; $r_{1,pre-post} =$ correlation between pre-test and post-test scores in the first group; $r_{2,pre-post} =$ correlation between pre-test and post-test scores in the second group; $\lambda_i =$ standardized factor loading for item i in 1 to I; $\varepsilon_i =$ error variance for item i in 1 to I; $r_{xy} =$ correlation between scale scores and criterion; $r_{xx'} =$ reliability of scale scores; $r_{yy'} =$ reliability of the criterion; $r =$ observed validity coefficient; $r^u =$ unrestricted validity coefficient; $S_r^2 =$ variance of the observed validity coefficient

Example of a 2×2 table

	Outcome of interest	Other outcome	
Intervention group	f_1	f_3	n_1
Control group	f_0	f_2	n_0

the complete code necessary to reproduce all analyses. Additionally, authors may overlook the inclusion of metadata, hindering the comprehension of variables within the dataset. On top of this, the inadequate reporting of crucial study details, such as the research procedure, sample characteristics, instrument details, and research design, remains unresolved even with the availability of a publicly accessible dataset. In essence, having access to a dataset does not guarantee that meta-analysts will acquire comprehensive information from the study necessary for inclusion in meta-analysis or meta-regression analyses, especially details suitable for moderator analyses. Hence, our recommendation is not only to provide access to the dataset and code used but also to adhere to the guidelines outlined in this manuscript.

When providing public access to the dataset and analytics code, it is crucial to consider specific key factors for ensuring the success of the process (see also Obels et al., 2020; Wilkinson et al., 2016). First, ensure the public accessibility and proper functionality of the website link hosting the documents. Second, provide a comprehensive codebook that clearly explains the coding for each variable. Third, include explanatory comments in the analytical code to guide fellow researchers through its execution. Finally, to overcome interoperability challenges and to ensure compatibility across different statistical software packages and versions, store data in universally readable formats such as .ASCII, .CSV, and .TXT. For comprehensive guidance on the process of data sharing, please refer to the step-by-step guide provided by Logan et al. (2021). This resource offers detailed insights and instructions to help one effectively navigate the various stages of sharing data.

Univariate statistics of the whole sample

Descriptive summary statistics (e.g., sample sizes, means, standard deviations, frequencies, and proportions) are crucial for accurately describing the variables under study and for calculating the most relevant effect sizes, including standardized mean differences, risk ratios, and odds ratios (see Table 1). It is important to highlight the necessity of providing this information for the final sample of participants after excluding dropouts. In longitudinal studies, providing descriptive statistics for each time point is particularly vital, especially in instances where participants were absent, or data were missing.

When studying qualitative categorical variables, such as dichotomous, nominal, or ordinal variables (e.g., socioeconomic status, type of stimuli, or type of task), frequencies and proportions should be reported for each category of the qualitative variable, regardless of whether it is an independent or dependent variable. For instance, in studies on inattentional blindness—where individuals may fail to notice unexpected

stimuli in their visual field due to focused attention on a different task or stimulus—the typical dependent variable is whether individuals notice an object unexpectedly introduced by the researcher in the task (e.g., Wiemer, et al., 2013), and authors should report the number and proportion of individuals who noticed the unexpected objects and those who did not.

Moving on to quantitative variables (e.g., age, income, or test scores), the descriptive statistics to be reported are means and standard deviations.[1] For instance, Harris (2004) examined the relationship between intelligence, achievement, openness to experience, and creativity. All these variables were quantitatively measured, and their means and standard deviations are appropriately presented in a table. Harris (2004) did not specify whether there was missing data, leading to the assumption that all variables are based on the complete sample. Ideally, it should be explicitly mentioned that no data were missing, or the sample size for each variable could have been indicated. Another instance is the study conducted by Goecke et al. (2020), where they investigated conflicting assertions regarding the overclaiming phenomenon (i.e., the inclination of individuals to overrate both their general cognitive abilities and their specific knowledge). The researchers measured various quantitative variables, including overclaiming, self-reported knowledge, and crystallized intelligence, and detailed their means, standard deviations, and corresponding sample sizes in a table. Notably, they provided precise information about the sample for each variable, with slight variations in sample sizes due to missing data. This meticulous reporting enables a future meta-analyst to discern the exact sample for each of these measures.

Descriptive statistics for the relationship between variables

When examining the relationship between variables, it is important to report the descriptive information associated with this relationship because this is the information commonly used by meta-analysts to calculate effect sizes. In the following subsections, we disaggregate this information by the types of variables involved in the relationship.

Relationship between categorical variables

The numerical information required for studying the relationship between categorical variables depends on the type of categorical variables under investigation. When studying

---

[1] Common analyses (e.g., regression analyses) assume that the residual scores on the outcome variable are normally distributed. A violation of this assumption or the existence of outliers may make means and standard deviations less informative. Therefore, information on the distribution of the scores is also required.

the relationship between two dichotomous or nominal variables, it is imperative to present a cross-tabulation with disaggregated frequencies. Such cross-tabulation provides the necessary information to calculate effect sizes, such as odds ratios and risk ratios, which are commonly used in meta-analyses of categorical data. For instance, consider a study investigating the association between smoking status (smoker vs. non-smoker) and the presence of lung cancer (yes vs. no). A cross-tabulation of these variables would display the frequencies of individuals falling into each combination of categories, for instance, the number of smokers diagnosed with lung cancer, non-smokers diagnosed with lung cancer, smokers not diagnosed with lung cancer, and non-smokers not diagnosed with lung cancer (see, for example, Morabia & Wynder, 1991). This detailed breakdown is essential for meta-analysts aiming to synthesize the association between these two variables across studies.

When studying the relationship between a dichotomous or a nominal variable and an ordinal variable or between two ordinal variables, it is crucial for researchers to provide access to the dataset containing raw data. In other words, if researchers utilize ordinal variables and aim for their study to be eligible for future meta-analyses, adherence to open science practices is imperative. This is because most effect sizes applicable to ordinal variables cannot be computed solely from descriptive summary statistics. For instance, to assess the magnitude of the difference between two groups in an ordinal variable, one might calculate the delta Cliff (Cliff, 1993), but raw data are indispensable (see Macbeth et al., 2011). Similarly, the correlation between two ordinal variables can be determined using Spearman or Kendall's tau-square correlation (Kendall, 1938), but once again, raw data are necessary for computation, as it involves examining concordant and discordant pairs of observations. Consequently, meta-analysts interested in effect sizes related to ordinal variables can include a particular study in their research synthesis only if the exact effect size of interest is reported or if authors have made their datasets publicly available.

Relationship between categorical and quantitative variables

When investigating the relationship between a categorical variable and a quantitative variable, means and standard deviations of the quantitative variable should be reported for each category of the categorical variable. Harris' (2004) study provides an example of how descriptive statistics for quantitative dependent variables are reported by pertinent groups. In this investigation, gender differences were examined, and a breakdown of means and standard deviations segregated by gender is provided in a table. This detailed presentation of descriptive statistics for relevant subgroups, such as based on gender, aids future meta-analysts in computing standardized mean

differences between genders across all measured variables. Especially in studies where the primary analysis involves an analysis of variance (ANOVA), it is crucial to report means, standard deviations, and sample sizes for each combination of categories of the qualitative variables used as the independent variable in the analyses. For instance, consider a two-factor ANOVA with independent variables such as socioeconomic status (low, medium, and high) and educational level (primary, high school, and university). In this case, means, standard deviations, and sample sizes should be reported for each of the $3 \times 3 = 9$ subgroups resulting from the combination of categories. This detailed reporting is essential as it enables meta-analysts to calculate standardized mean differences for any of the resulting subgroups. It is important to note that this descriptive information should be reported regardless of the primary researcher's specific focus, which typically revolves around the interaction between the independent variables, and it does not necessarily have to be included in the main text; it can be relocated to the supplementary materials.

Relationship between quantitative variables

Pearson correlation coefficients summarize the (linear) relationship between two quantitative variables. These coefficients are incredibly useful in meta-analysis for several reasons. Firstly, correlation coefficients serve as effect sizes that can be readily integrated into meta-analytic datasets. Second, many partial effect sizes can be calculated from correlation coefficients, such as partial- and semi-partial correlations and standardized regression coefficients (Aloe & Becker, 2009, 2012; Becker, 1992; Fernández-Castilla et al., 2019). However, if researchers only report the results of multiple regression models (i.e., unstandardized or standardized regression coefficients), correlation coefficients cannot be back-calculated,[2] and this is a reason why many primary studies are often discarded for meta-analysis. Although a procedure to convert regression coefficients to correlations has been proposed (Peterson & Brown, 2005), it does not work correctly in many scenarios (Aloe, 2015). Hence, simply reporting correlations among quantitative variables enables the calculation of many effect sizes that might be of interest to meta-analysts.

A third reason why correlation coefficients should always be reported is that, to implement multivariate meta-analytic models, the correlation between the raw scores of the variables of interest is needed. For instance, imagine that a meta-analyst is interested in synthesizing standardized mean differences that reflect the effectiveness of a given psychological intervention in reducing both anxiety and depressive symptoms, and that most studies report these two results.

---

[2] An exception would be the method presented by Fernández-Castilla et al. (2019), where correlation coefficients can be estimated from standardized regression coefficients under some scenarios.

Since there are two correlated dependent variables within studies (anxiety and depression), a multivariate meta-analysis would have to be carried out to synthesize these effect sizes (Becker, 2000; Kalaian & Raudenbush 1996). To conduct this type of analysis, the covariance between the standardized mean differences reported in the same study (presumably one for depression and one for anxiety) needs to be estimated in advance (see Hedges & Olkin, 1985), and to calculate it, information on the correlation between the raw depression and anxiety scores is needed. By reporting the correlation coefficients between all quantitative variables, future meta-analysts will be able to retrieve this information to apply more sophisticated statistical methods, eventually leading to more precise meta-analytic estimates.

A final reason why reporting correlation coefficients is important is that new methods have been developed in the field of meta-analysis, such as meta-analytic structural equation modeling (also known as MASEM, Cheung, 2015; Jak, 2015; or one-stage MASEM [OSMASEM], Jak & Cheung, 2020). This methodology allows one to perform meta-analysis of more complex structural equation models, including mediation models (e.g., Ng et al., 2023), path analyses (Smith et al., 2022), or confirmatory factor analyses (Said-Metwaly et al., 2018). The input required for conducting MASEM is the correlations between the variables of interest organized in a correlation matrix. By reporting all the possible correlations of one's dataset in a correlation matrix, meta-analysts performing MASEM could easily include all the correlations between their variables of interest.

Intraclass correlation coefficient and variance estimates in cluster-randomized studies

In primary research within the realms of psychology and educational sciences, it is commonplace to encounter hierarchical structures wherein observations are nested within higher-level clusters. Examples include students nested within classrooms or observations nested within participants in repeated measures designs. This hierarchical structure necessitates consideration not only during data analysis but also in the calculation of certain effect sizes, such as the standardized mean difference (Hedges, 2007; Snijders, 2005).

Consider, for instance, a scenario where two groups of participants from distinct experimental conditions are compared (level 1), and these participants are further grouped into different centers, forming the cluster at level 2. When calculating the standardized mean difference that compares means across these experimental conditions, it becomes imperative to acknowledge that participants are nested within different clusters (centers in this case). Consequently, participants belonging to the same center are expected to exhibit greater similarity than those from different centers.

There is no singular formula for calculating a standardized mean difference for clustered designs. The mean difference between groups may be standardized by the square root of the pooled within-cluster variance, the between-clusters variance, or the total variance, representing the sum of the two variances. Therefore, to facilitate the calculation of any of these versions of effect sizes, a meta-analyst must possess information on (1) the mean of the two compared groups, (2) the between-clusters variance, and (3) the within-cluster variance. These sources of variability can also be estimated from each other if the intraclass correlation coefficient is available. This coefficient signifies the correlation between observations within the same cluster, and the relevant formulas can be found in Borenstein and Hedges (2019). The intraclass correlation coefficient, coupled with the total sample size and average cluster size, is also essential for calculating the sampling variances of these effect sizes. Consequently, it is of utmost importance to thoroughly report all this information in studies employing such designs.

Pearson correlations between repeated measures

In meta-analysis, it is often of interest to include data from matched group experimental designs meant to test the effectiveness of an independent variable (e.g., intervention, program, or experimental condition). Typically, in each study, standardized mean differences for repeated measures (see $g_{rm(1)}$ in Table 1) or standardized mean changes (see $g_{igpp(1)}$ in Table 1) are calculated for posterior synthesis. Importantly, the formulas for these effect sizes incorporate the correlation between pre- and post-measures. Specifically, this correlation is essential for determining the standard deviation of the difference (S_{within}), which serves as the denominator in the formula for computing the standardized mean difference for repeated-measures designs ($g_{rm(1)}$ in Table 1). Furthermore, this correlation between pre- and post-test scores is necessary for calculating the sampling variance of this effect size (see Morris and DeShon, 2002). Similarly, to calculate a standardized mean change (i.e., the standardized difference in the extent of change within one group relative to the change observed in another group, see $g_{igpp(1)}$ in Table 1), the correlation between pre- and post-measures within each involved group is also required.

Since authors seldom report this correlation, formulas have been proposed to circumvent its inclusion in the calculation of these effect sizes (see, for instance, Becker, 1988; see formula for $g_{rm(2)}$ and $g_{igpp(2)}$ in Table 1). However, this pre/post score correlation is still essential for computing the sampling variance of these effect sizes (see $S^2_{g_{rm(2)}}$ and $S^2_{g_{igpp(2)}}$ in Table 1). Therefore, when utilizing standardized mean differences or standardized mean changes in meta-analysis, the correlation between pre- and post-measures often needs

to be estimated or imputed. Hence, we strongly encourage primary researchers to incorporate this correlation in their reports, along with any other pertinent descriptive information.

Reliability of the measurements

Reliability is commonly defined as the proportion of true score variance to total score variance (Novick, 1966). Reliability coefficients provide information on the precision of scores from psychological assessments. In psychological science, measurements frequently contain non-negligible degrees of error. For instance, self-reported outcomes may include nuisance related to the distortions in individuals' self-perception or understanding of the response scale. These measurement errors are generally assumed to be random variations that cause scores to deviate from their true values.

Although often disregarded, the results of a primary study containing psychological assessments are largely influenced by measurement reliability. As an example, in the relationship between general intelligence and job performance, if both measures are precise (e.g., obtained using long questionnaires), the estimated regression/correlation coefficients are likely to approximate the true relationship between these constructs. On the other hand, if measurement reliability is low (e.g., using fewer or more imprecise questions), the coefficients between variables may be largely underestimated. To illustrate, a correlation of 0.51 between intelligence and job performance (e.g., Schmidt & Hunter, 2004) could be substantially reduced, to approximately 0.36, if both measurements have reliability coefficients of around 0.70.

Meta-analytic studies are often aimed at summarizing generalized coefficients for the relationships between constructs beyond one specific sample. Correcting these underestimated regression/correlation coefficients relies on the reliability indices reported in primary studies. Authors are encouraged to report reliability coefficients (e.g., Cronbach's α or McDonald's ω) of their measurements. Finally, it is important to note that the reliability reported in an instrument's manual or in the original validation studies may not precisely match reliability in empirical studies. Due to range restrictions of the scores and additional noise due to various random factors, the reliability in an empirical study can differ from the one in the original validation study. Hence, authors of primary studies are encouraged to report the reliability of measurements obtained in their datasets. This not only enhances their reporting but also makes their studies eligible for future reliability generalization meta-analysis.

Negative results

Researchers may conduct a study and find an effect that is either statistically nonsignificant or contradicts a hypothesis, referred to as a negative finding. Negative findings face a greater publication challenge than their positive counterparts (Fanelli, 2010; Franco et al., 2014). Researchers may fuel publication bias by selectively reporting positive findings or refraining from submitting studies with negative findings. This behavior is often driven by the anticipation of low acceptance rates, or the fear of professional consequences associated with publishing findings that challenge well-confirmed hypotheses or theories (Shields, 2000; Therrien & Cook, 2018). Journal editors and reviewers may also contribute to publication bias by rejecting submissions with null findings.

Publication bias has been observed in various fields, including medicine, social sciences, and psychology, indicating a widespread phenomenon (Therrien & Cook, 2018). Publication bias may inflate the estimates of relationships between variables and treatment effects in meta-analyses. The inclusion of even a few unpublished findings could substantially influence conclusions drawn from the literature (Howard et al., 2009; Polanin et al., 2016). Publication bias distorts scientific literature, leading to the formulation of hypotheses or taking decisions in practice based on inaccurate information, wasting research opportunities and funding and violating an implicit contract with funders (Shields, 2000). Moreover, when negative findings go unpublished, researchers may expend resources conducting studies that have already proven unsuccessful (Fanelli, 2012). The potential bias in the literature, however, is not the only problem with not reporting findings. We also have an ethical responsibility to our study participants who invest their valuable time and resources, trusting that their contributions benefit others. Failure to publish study findings violates this trust and may be deemed scientific misconduct (Chalmers & Moher, 1993; Shields, 2000). Additionally, we owe transparency to donors and taxpayers who support our research.

To encourage the publication of negative findings, it is crucial to recognize the value of negative results on par with positive ones. Acknowledging that understanding the absence of an effect holds equal importance to identifying its presence is essential (Fox & Kaufman, 2018). Instead of planning studies solely to determine "what works," a shift to planning studies to understand "how to make things work better" allows for useful insights from positive or negative findings (Jacob et al., 2019). By shifting our perspective and acknowledging the importance of negative findings, we contribute to a more balanced and comprehensive scientific literature, fostering a culture that appreciates the diverse outcomes of rigorous research efforts.

Recognizing their significance, initiatives have been undertaken to improve the visibility of negative findings in scientific literature through diverse approaches. For instance, certain journals have been initiated exclusively dedicated to publishing negative findings, such as the *Journal of Negative Results*, *Journal of Negative Results in BioMedicine*,

Journal of Pharmaceutical Negative Results, *Nature's* Negative Results section, and *Positively Negative (PLOS One)*. In addition, mainstream journals have allocated special issues specifically for null findings (see, for instance, Landis et al., 2014; Therrien & Cook, 2018). However, this approach may inadvertently introduce bias favoring negative outcomes (Mlinarić et al., 2017). Publishing criteria should thus prioritize study quality and statistical power, irrespective of the direction and significance of the results.

Journal editors and reviewers can also play a pivotal role in shaping positive attitudes and behaviors regarding negative findings. For instance, editors can explicitly express in the author guidelines the openness to publish well-designed studies with null findings (Hubbard & Armstrong, 1992). Editors can also promote or mandate registered reports, where study plans are submitted for pre-publication review based on research design. If accepted, the study is published regardless of the reported findings, minimizing the likelihood of result-driven deviations or studies being overlooked in file drawers (Cook & Therrien, 2017). Moreover, during the revision process, editors and reviewers commonly ask for the removal of information that is deemed nonessential, frequently tied to negative findings. While brevity is important, we should not sacrifice information. Unless entirely unrelated to the primary research question, it is advisable to report findings regardless of their direction, even if placed in supplementary material—thus, providing more information is generally preferable (Landis et al., 2014). Such practices could address publication bias by directly publishing more studies with negative findings and indirectly affirming their value and publishability, encouraging researchers to submit rather than keep them in a file drawer (Cook & Therrien, 2017).

Study Eligibility for Meta-Analysis Inclusion (SEMI) checklist

Many reporting guidelines have been provided for studies of different fields: the STROMS checklist for research on human microbiome (Mirzayi, et al., 2021), the AGREE Reporting Checklist for clinical research (Brouwers et al., 2016), and the CROSS checklist for survey studies (Sharma et al., 2021). In this direction, some interesting initiatives have emerged, such as the EQUATOR Network (Altman et al. 2008), which brings together different resources and checklists that aim to improve the accuracy of the reporting and the quality of the research. There are also well-known reporting guidelines developed to properly report information in meta-analyses and systematic reviews (the PRISMA statement, Page et al., 2021; the REGEMA checklist, Sánchez-Meca et al., 2021 in reliability generalization meta-analyses). However, there is currently a lack of reporting guidelines specifically aimed at enhancing the

odds of a study being retrieved and being eligible for a meta-analysis, and that is the gap aimed to be filled with this manuscript.

In this section, we provide the SEMI checklist to supplement extant reporting guidelines in the hope of improving the completeness of information in primary empirical reports and thus optimizing for inclusion in future meta-analyses. The SEMI checklist may be used in conjunction with other checklists assessing basic reporting prerequisites (e.g., PRISMA, STROBE, and CONSORT), aiming to maximize the quality of reporting practices and facilitate accumulated meta-analytic knowledge.

The development of the checklist items was informed by existing reporting guidelines, our own experience in meta-analysis research, and consultation with expert researchers in the field. It was also guided by the PICO model, which is frequently used for planning literature search and study selection in research synthesis (McGowan et al., 2016).

Initially, The SEMI checklist involved 30 items, followed by a "yes/no/not applicable" judgment, covering five key parts of a paper: title and abstract, background, methods, results, and open science. We incorporated items related to study title and abstract to encourage researchers to consider reporting information that support meta-analysts to retrieve the study in database searching and to conduct title/abstract screening. We also incorporated items related to a study background to facilitate locating relevant studies via backward reference searching. In the Methods section, we present elements pertaining to the accurate reporting of study characteristics, crucial for the subsequent execution of moderator analyses in meta-analysis. Next, in the Results section, our focus is on elements related to the proper reporting of numerical information, essential for calculating effect sizes. We also incorporated items to prompt researchers to report results in a sequence that mirrors the description of analyses outlined in the Methods section and to ensure coherence between the textual results with those displayed in the tables and figures. This can help mitigate ambiguity and potential misinterpretation, offering meta-analysts a clear roadmap to navigate the study's design, methods, and results without unnecessary confusion. Finally, we also include some items related to open science practices. This initial version of the checklist underwent review by four external methodological experts in the field of meta-analysis, some with more than two decades of experience in meta-analysis, who provided valuable feedback to refine the tool. Incorporating expert opinions, we revised existing items and introduced new ones, resulting in a final set of 28 items.

We make the SEMI available in Table 2 for the research community and will register it on the EQUATOR website to enhance dissemination. We recommend journals and publishers endorse the use of the SEMI by referring to it in their instructions to authors and consider utilizing it in their review process.

Table 2 Study eligibility for meta-analysis inclusion checklist

Item	Y	N	NA

Title and abstract

1. The key concepts, constructs, and variables under investigation are clearly mentioned in the title and/or abstract.

2. The abstract gives relevant details about the study objectives, methods, and results.

Background

3. Relevant literature (including reviews and meta-analyses) is summarized and clearly cited.

Methods

4. The sample size, including that of the entire sample and each subsample, is reported. The number of missing values is given for each variable, and the sample size used for each analysis is reported. In the case of longitudinal studies, the sample size at each time point is reported.

5. Statistics describing participant characteristics (e.g., proportion identifying as men, mean age, proportion of sample by race/ethnicity), study context and procedures that may (substantially) influence the studied effects are reported.

6. Other publications based on the same data, or a portion thereof, are clearly cited.

7. There is a description of how each variable is operationally defined and measured.

8. Details of how the measurement tools are administered and scored are provided, together with a measure of reliability on the current sample.

9. Details of the type of study design (e.g., correlational, comparative, or experimental) are provided, possibly together with a bibliographic source for further details.

10. Details of the study procedures are provided, including where, how, and when data are collected.

11. There is a description of how data categories are defined or how continuous variables are categorized. When reporting data from a subsample, details on subsample descriptions and selection criteria subsample are provided.

12. Details of the data-analytic methods used are provided (e.g., statistical tests, model fitted, estimation procedure, software, options chosen, significance level, whether the test is two-sided or one-sided, degrees of freedom, how cluster data are handled if needed, and whether missing data imputation methods were used and which ones).

13. A risk of bias assessment tool is consulted to ensure the inclusion of all methodological details required for evaluating the study's risk of bias.

Results

14. For categorical variables, frequencies of all categories are reported for the final sample and relevant subgroups, after removing dropouts. When studying the association between categorical variables, a cross-tabulation is provided with disaggregated frequencies.

15. For quantitative variables, means and standard deviations are provided for the whole sample and relevant subgroups.

16. For nested data structures, information on the intraclass correlation coefficient, the between-clusters variance, and the pooled within-cluster variance is reported.

17. The correlation matrix between all quantitative variables under investigation is reported. When missing data are imputed, correlations based on the original incomplete data are provided.

18. For longitudinal studies, the timing of measurements and the correlation between subsequent measures is reported, also for any relevant subgroup.

19. Test statistics and associated p-values (and degrees of freedom where relevant) are reported, also for negative findings.

20. Effect sizes related to the research questions are presented along with (references to) the corresponding formulas used for their calculation.

21. The results are reported in sufficient detail and clarity, following the description of the analyses in the methods-section (e.g., following the same order).

22. The results presented in the text align with those depicted in the tables and figures.

23. Tables and figures are appropriately labelled, understandable and referred to in the text.

Open science practices

24. A statement indicating the availability and location of raw study data (and if applied of the protocol or registered report) is provided.

25. If a protocol or registered report was developed before the investigation, it is clarified how the investigation deviates from the initial planning.

26. A code book explaining the variables in the dataset is provided.

27. Relevant codes/syntax that reproduce the analyses are provided.

28. Additional information or materials that could enhance understanding of methods or results are included in appendices or supplementary materials.

Y = Yes; N = No; NA = Not applicable

Discussion

Meta-analysis has emerged as a powerful tool for consolidating scientific knowledge and informing decision-making. However, the accurate execution of various stages of a meta-analysis may be hindered by the inaccurate reporting of information in primary research studies. If studies cannot be found or if effect sizes cannot be computed, they will be excluded from the research synthesis, ultimately impacting the statistical power to detect a significant overall effect or even inducing bias. Likewise, if the characteristics of the studies cannot be effectively encoded, there will be missing information in the moderator analysis, which in turn will affect the analytical power (Pigott, 2019). Although imputation techniques exist to prevent this problem (e.g., Lee & Beretvas, 2022), no technique will yield the same accurate estimates as having all the data available for the analyses.

For this reason, we have introduced the SEMI checklist, which can be utilized to assess the suitability of a study for inclusion in future meta-analyses. To the best of our knowledge, this checklist represents one of the first endeavors to improve the reporting quality of primary studies, with a specific focus on their potential inclusion in a meta-analysis. In a similar vein, Chow et al. (2023) have offered valuable recommendations for reporting specific elements of studies, such as procedures, results, and open access practices. Our checklist broadens the scope of Chow's checklist to include additional critical elements. This encompasses aspects such as the study's title, abstract, background, sample characteristics, and other results essential for calculating various effect sizes in meta-analysis, thereby ensuring a more comprehensive reporting framework.

Hopefully, the use of the SEMI checklist and the Chow et al. (2023) guidelines can assist authors in describing the conducted research in sufficient detail, assist editors and reviewers in evaluating the comprehensiveness of reports submitted for publication, and ultimately maximize the use of research results in the quantitative synthesis. We believe that adhering to the suggested checklist can substantially enhance the reporting standard of primary studies. This, in turn, will ultimately contribute to conducting more precise and reliable meta-analyses.

References

Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic review. *Review of Educational Research, 90*, 6–23. https://doi.org/10.3102/0034654319854352

Aloe, A. M. (2015). Inaccuracy of regression results in replacing bivariate correlations. *Research Synthesis Methods, 6*, 21–27. https://doi.org/10.1002/jrsm.1126

Aloe, A. M., & Becker, B. J. (2009). Teacher verbal ability and school outcomes: Where is the evidence? *Educational Researcher, 38*(8), 612–624. https://doi.org/10.3102/0013189X09353939

Aloe, A. M., & Becker, B. J. (2012). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics, 37*, 278–297. https://doi.org/10.3102/1076998610396901

Altman, D. G., Simera, I., Hoey, J., Moher, D., & Schulz, K. (2008). EQUATOR: Reporting guidelines for health research. *The Lancet, 371*(9619), 1149–1150. https://doi.org/10.1016/S0140-6736(08)60505-X

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*, 3–25. https://doi.org/10.1037/amp0000191

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational and Behavioral Statistics, 17*, 341–362. https://doi.org/10.3102/10769986017004341

Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & E. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 499–525). Academic Press.

Booth, A. (2006). Clear and present questions: Formulating questions for evidence-based practice. *Library Hi Tech, 24*, 355–368. https://doi.org/10.1108/07378830610692127

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of Research Synthesis and Meta-Analysis* (3rd ed., pp. 207–243). Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

Brouwers, M. C., Kerkvliet, K., Spithoff, K., & AGREE Next Steps Consortium. (2016). The AGREE Reporting checklist: A tool to improve reporting of clinical practice guidelines. *British Medical Journal*, i1152. https://doi.org/10.1136/bmj.i1152

Chalmers, I., & Moher, D. (1993). Publication bias. *Lancet, 342*, 1116. https://doi.org/10.1016/0140-6736(93)92099-F

Cheung, M.W.-L. (2015). *Meta-Analysis: A structural equation modeling approach*. John Wiley & Sons.

Chow, J. C., Sandbank, M., & Hampton, L. H. (2023). Guidance for increasing primary study inclusion and the usability of data in meta-analysis: A reporting tutorial. *Journal of Speech, Language, and Hearing Research*, 1-9. https://doi.org/10.1044/2023_JSLHR-22-00318

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin, 114*, 494–509. https://doi.org/10.1037/0033-2909.114.3.494

Cook, B. G., & Therrien, W. J. (2017). Null effects and publication bias in special education research. *Behavioral Disorders, 42*, 149–158. https://doi.org/10.1177/0198742917709473

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research, 22*, 1435–1443. https://doi.org/10.1177/1049732312452938

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation. https://doi.org/10.7758/9781610448864

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. https://doi.org/10.1007/BF02310555

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2008). Analyzing data and undertaking meta-analyses. In I. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 243–296). John Wiley & Sons.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*, 917–928. https://doi.org/10.1093/jpepsy/jsp004

Egger, M., & Smith, G. D. (1997). Meta-analysis: Potentials and promise. *British Medical Journal, 315*, 1371–1374. https://doi.org/10.1136/bmj.315.7119.1371

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PloS One, 5*, e10068. https://doi.org/10.1371/journal.pone.0010068

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

Fernández-Castilla, B., Aloe, A. M., Declercq, L., Jamshidi, L., Onghena, P., Natasha Beretvas, S., & Van den Noortgate, W. (2019). Concealed correlations meta-analysis: A new method for synthesizing standardized regression coefficients. *Behavior Research Methods, 51*, 316–331. https://doi.org/10.3758/s13428-018-1123-7

Fox, M. P., & Kaufman, J. S. (2018). The WelTel Trial in context and the importance of null findings. *The Lancet Public Health, 3*, e107–e108. https://doi.org/10.1016/S2468-2667(18)30004-5

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*, 1502–1505. https://doi.org/10.1126/science.1255484

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

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis for social research*. Sage.

Goecke, B., Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). Testing competing claims about overclaiming. *Intelligence, 81*, 101470. https://doi.org/10.1016/j.intell.2020.101470

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Lawrence Erlbaum Associates Publishers.

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia, 14*, 29–37.

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, 5(8), 180448. https://doi.org/10.1098/rsos.180448

Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal Psychological Science: An observational study. *Royal Society Open Science, 8*(1), 201494. https://doi.org/10.1098/rsos.201494

Harris, J. A. (2004). Measured intelligence, achievement, openness to experience, and creativity. *Personality and Individual Differences, 36*(4), 913–929. https://doi.org/10.1016/S0191-8869(03)00161-2

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341–370. https://doi.org/10.3102/1076998606298043

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

Howard, G. S., Lau, M. Y., Maxwell, S. E., Venter, A., Lundy, R., & Sweeny, R. M. (2009). Do research literatures give correct answers? *Review of General Psychology, 13*, 116–121. https://doi.org/10.1037/a0015468

Hubbard, R., & Armstrong, J. S. (1992). Are null results becoming an endangered species in marketing? *Marketing Letters, 3*, 127–136. https://doi.org/10.1007/BF00993992

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*, 594–612. https://doi.org/10.1037/0021-9010.91.3.594

Jacob, R. T., Doolittle, F., Kemple, J., & Somers, M. A. (2019). A framework for learning from null results. *Educational Researcher, 48*, 580–589. https://doi.org/10.3102/0013189X19891955

Jak, S. (2015). *Meta-analytic structural equation modelling*. Springer. https://doi.org/10.1007/978-3-319-27174-3

Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods, 25*, 430. https://doi.org/10.1037/met0000245

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1*, 227–235. https://doi.org/10.1037/1082-989X.1.3.227

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*(2), 137–151. https://doi.org/10.1037/a0028086

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*, 81–93. https://doi.org/10.1093/biomet/30.1-2.81

Krumholz, H. M. (2012). Open science and data sharing in clinical research: Basing informed decisions on the totality of the evidence. *Circulation: Cardiovascular Quality and Outcomes, 5*, 141–142. https://doi.org/10.1161/circoutcomes.112.965848

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Landis, R. S., James, L. R., Lance, C. E., Pierce, C. A., & Rogelberg, S. G. (2014). When is nothing something? Editorial for the null results special issue of Journal of Business and Psychology. *Journal of Business and Psychology, 29*, 163–167. https://doi.org/10.1007/s10869-014-9347-8

Lee, J., & Beretvas, S. N. (2022). Comparing methods for handling missing covariates in meta-regression. *Research Synthesis Methods, 14*, 117–136. https://doi.org/10.1002/jrsm.1585

Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data sharing in education science. *AERA Open, 7*, 233285842110064. https://doi.org/10.1177/23328584211006475

Macbeth, G., Razumiejczyk, E., & Ledesma, R. D. (2011). Cliff´s Delta Calculator: Un programa para calcular el tamaño del efecto no paramétrico al comparar dos grupos de observaciones. *Universitas Psychologica, 10*, 545–555.

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology, 75*, 40–46. https://doi.org/10.1016/j.jclinepi.2016.01.021

Mirzayi, C., et al. (2021). Reporting guidelines for human microbiome research: The STORMS checklist. *Nature Medicine, 27*, 1885–1892. https://doi.org/10.1038/s41591-021-01552-x

Mlinarić, A., Horvat, M., & ŠupakSmolčić, V. (2017). Dealing with the positive publication bias: Why you should really publish your negative results. *Biochemia Medica, 27*, 447–452. https://doi.org/10.11613/BM.2017.030201

Morabia, A., & Wynder, E. L. (1991). Cigarette smoking and lung cancer cell types. *Cancer, 68*, 2074–2078. https://doi.org/10.1002/1097-0142(19911101)68:9%3C2074::AID-CNCR2820680939%3E3.0.CO;2-X

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological methods, 7*, 105–125. https://doi.org/10.1037//1082-989x.7.1.105

Ng, M. Y., DiVasto, K. A., Cootner, S., Lipsey, M. W., & Weisz, J. R. (2023). How do cognitive behavioral therapy and interpersonal psychotherapy improve youth depression? Applying meta-analytic structural equation modeling to three decades of randomized trials. *Psychological Bulletin, 149*, 507–548. https://doi.org/10.1037/bul0000395

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*, 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science, 3*, 229–237. https://doi.org/10.31234/osf.io/fk8vh

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241–286. https://doi.org/10.1006/ceps.2000.1040

Özsoy, G., & Özsoy, S. (2013). Effect size reporting in educational research. *Elementary Education Online, 12*, 334–346. http://earsiv.odu.edu.tr:8080/xmlui/handle/11489/1756. Accessed 1 Jan 2024.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods, 23*, 208–225. https://doi.org/10.1037/met0000126

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*, 175–181. https://doi.org/10.1037/0021-9010.90.1.175

Pigott, T. D. (2019). Missing data in meta-analysis. In Russell Sage (Ed.), *Handbook of Research Synthesis* (pp. 367–381). Russell Sage Foundation.

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research, 90*, 24–46. https://doi.org/10.3102/0034654319877153

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86*, 207–236. https://doi.org/10.3102/0034654315582067

Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods, 10*, 330–342. https://doi.org/10.1002/jrsm.1354

Ray, J. W., & Shadish, W. R. (1996). How interchangeable are different estimators of effect size? *Journal of Consulting and Clinical Psychology, 64*, 1316–1325. https://doi.org/10.1037/0022-006X.64.6.1316

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *British Medical Journal, 340*, c221. https://doi.org/10.1136/bmj.c221

Rytwinski, T., Cooke, S. J., Taylor, J. J., Roche, D. G., Smith, P. A., Mitchell, G. W., Smokorowski, K. E., Prior, K. A., & Bennett, J. R. (2021). Acting in the face of evidentiary ambiguity, bias, and absence arising from systematic reviews in applied environmental science. *Science of the Total Environment, 775*, 145122. https://doi.org/10.1016/j.scitotenv.2021.145122

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van den Noortgate, W. (2018). The factor structure of the Figural Torrance Tests of Creative Thinking: A meta-confirmatory factor analysis. *Creativity Research Journal, 30*, 352–360.

Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods, 12*, 516-536. https://doi.org/10.1002/jrsm.1487

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis* (3rd ed.). Sage.

Sharma, A., Minh Duc, N. T., Luu Lam Thang, T., Nam, N. H., Ng, S. J., Abbas, K. S., & Karamouzian, M. (2021). A consensus-based checklist for reporting of survey studies (CROSS). *Journal of General Internal Medicine, 36*, 3179-3187 https://doi.org/10.1007/s11606-021-06737-1

Shields, P. G. (2000). Publication bias is a scientific problem with adverse ethical outcomes: The case for a section for null results. *Cancer Epidemiology Biomarkers & Prevention, 9*, 771–772.

Smith, M. M., Hewitt, P. L., Sherry, S. B., Flett, G. L., & Ray, C. (2022). Parenting behaviors and trait perfectionism: A meta-analytic test of the social expectations and social learning models. *Journal of Research in Personality, 96*, 104180. https://doi.org/10.1016/j.jrp.2021.104180

Snijders, T. A. (2005). Power and sample size in multilevel modeling. *Encyclopedia of Statistics in Behavioral Science, 3*, 1573. https://doi.org/10.1017/cbo9780511790942.026

Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., & Higgins, J. P. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ, 366*. https://doi.org/10.1136/bmj.l4898

Therrien, W. J., & Cook, B. G. (2018). Introduction to special issue: Null effects and publication bias in learning disabilities research. *Learning Disabilities Research & Practice, 33*, 5–10. https://doi.org/10.1111/ldrp.12163

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019). A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods, 10*, 161–179. https://doi.org/10.1002/jrsm.1338

Trusty, J., Thompson, B., & Petrocelli, J. V. (2004). Practical guide for reporting effect size in quantitative research in the Journal of Counseling & Development. *Journal of Counseling & Development, 82*, 107–110. https://doi.org/10.1002/j.1556-6678.2004.tb00291.x

Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. *Cleveland Clinic Journal of Medicine, 75*(6), 431–439. https://doi.org/10.3949/ccjm.75.6.431

Whiting, P., Wolff, R., Savović, J., Mallett, S., & Devine, B. (2023). LATITUDES network. Retrieved from https://www.latitudes-network.org/. Accessed 1 Jan 2024.

Wiemer, J., Gerdes, A. B. M., & Pauli, P. (2013). The effects of an unexpected spider stimulus on skin conductance responses and eye movements: An inattentional blindness study.

Psychological Research, 77, 155–166. https://doi.org/10.1007/s00426-011-0407-7

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Bouwman, J. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 109. https://doi.org/10.1038/sdata.2016.18

Wood, K., & Simons, D. J. (2017). Selective attention in inattentional blindness: Selection is specific, but suppression is not. *Collabra Psychology, 3*, 19. https://doi.org/10.1525/collabra.90