



HHS Public Access

Author manuscript

Appl Psychol Health Well Being. Author manuscript; available in PMC 2020 November 01.

Published in final edited form as:

Appl Psychol Health Well Being. 2019 November ; 11(3): 353–381. doi:10.1111/aphw.12169.

Best Practice Guidelines and Essential Methodological Steps to Conduct Rigorous and Systematic Meta-Reviews

Emily Alden Hennessy¹, Blair T. Johnson¹, Ciara Keenan^{2,3}

¹Institute for Collaboration on Health, Intervention, and Policy (InCHIP). Department of Psychological Sciences. University of Connecticut. Storrs, CT, USA, 06269-1248 ²Centre for Evidence and Social Innovation, Queen's University Belfast. University Road, Belfast, BT7 1NN, Northern Ireland, United Kingdom ³Campbell UK and Ireland

Abstract

Background.—A growing body of primary study and systematic review literature evaluates interventions and phenomena in applied and health psychology. Reviews of reviews (i.e., meta-reviews) systematically synthesize and utilize this vast and often overwhelming literature; yet, currently there are few practical guidelines for meta-review authors to follow.

Objective.—The aim of this article is to provide an overview of the best practice guidelines for all research synthesis and to detail additional specific considerations and methodological details for the best practice of conducting a rigorous meta-review.

Methods.—This article provides readers with six systematic and practical steps along with accompanying examples to address with rigor the unique challenges that arise when authors familiar with systematic review methods begin a meta-review: (a) detailing a clear scope, (b) identifying synthesis literature through strategic searches, (c) considering datedness of the literature, (d) addressing overlap among included reviews, (e) choosing and applying review quality tools, and (f) appropriate options for handling the synthesis and reporting of the vast amount of data collected in a meta-review.

Conclusions.—We have curated best practice recommendations and practical tips for conducting a meta-review. We anticipate that assessments of meta-review quality will ultimately formalize best-method guidelines.

Keywords

Meta-review; Best-practice guidelines; Methods; Research synthesis

A simple Google Scholar search with the term, “review of reviews” returns over 16,000 records, approximately 2,200 (14%) of which appeared in the past year. However

Corresponding author: Emily Alden Hennessy, MPhil, PhD: emily.hennessy@uconn.edu. Institute for Collaboration on Health, Intervention, and Policy. Department of Psychological Sciences. University of Connecticut. 2006 Hillside Road, Unit 1248. Storrs, CT, USA, 06269-1248.

Conflict of Interest:

The authors have no conflicts of interest to disclose.

voluminous the prevalence of meta-reviews in the literature, to date there is no comprehensive source of information on best practices in conducting a review of reviews, hereafter labelled as a *meta-review*.¹ Accordingly, in this article, we pull together existing methodological literature on research synthesis and specifically, meta-reviews, to provide clear guidelines of issues to consider and resulting processes to undertake when conducting a meta-review in applied psychology. This article has three primary aims:

1. Describe the role(s) of meta-reviews and the primary reasons why they should be conducted;
2. Present appropriate options for addressing meta-review challenges; namely, conducting rigorous but precise searches, handling overlap, up-to-datedness, synthesis options, and risk of bias assessment;
3. Outline areas for advances in meta-review synthesis methods.

Ultimately, we envision that readers of this article will become more informed synthesis authors, peer reviewers, and consumers of systematic reviews.

Synthesis Literature and the Role of Meta-reviews

The accumulation of research across all scientific disciplines is staggering. For such evidence to be put to maximal use, meta-reviews are important and even necessary (Bastian, Glasziou, & Chalmers, 2010; Ioannidis, 2016). Meta-reviews, which pull together existing synthesis literature, can have tremendous influence on research, practice, and policy; indeed, if conducted appropriately, synthesis literature is considered the strongest level of evidence, with meta-reviews atop the evidence pyramid (Biondi-Zoccai, 2016). Meta-reviews serve at least two purposes: (a) they can help identify the strongest evidence base, such as by separating the low versus high quality synthesis research, or by examining how apples and oranges compare in order to better understand fruit (thereby combatting a common critique of synthesis literature), or (b) by providing an “‘umbrella’ that prevents you from getting ‘soaking wet’ under ‘a rain of evidence’” (Abbate in Biondi-Zoccai, 2016, p. viii). In short, at their best, meta-reviews sum up the existing evidence on a subject as reflected in extant reviews.

Clearly, meta-reviews’ benefits will be limited unless they are conducted with the highest synthesis standards, which to date have not been unified in a single source. Fortunately, rigorous approaches for conducting primary evidence syntheses (e.g., systematic review, meta-analysis) also apply to meta-reviews. These methods stress that the process should be transparent and reproducible: Science is cumulative, and scientists should cumulate scientifically. The process of cumulation (meta-analysis, evidence synthesis, meta-review) should use methods to reduce biases and the play of chance. Thus, the following steps should always be followed whether a meta-analysis or meta-review is being conducted, including: developing research questions and registering a plan for examining them, conducting a systematic search for literature, independent and duplicate literature screening

¹Note, the term “meta-review” is used intentionally versus other such commonly used terms, such as “overview” (which could refer to many things that are not systematic reviews) or “umbrella review” (which refer to synthesis of primary studies and review literature and encompass a much broader scope and further methodological considerations).

and data extraction, using appropriate synthesis methods, conducting and reporting a standardized assessment of potential bias, and transparent reporting of all steps. These processes and rationale are already well explicated in numerous texts (e.g., see Borenstein, Hedges, Higgins, & Rothstein, 2011; Lipsey & Wilson, 2001; The Cochrane Collaboration., 2011; Waddington et al., 2012) and will not be detailed here, nevertheless, there are additional considerations for conducting meta-reviews that should be addressed and that are the focus of this article.

Steps for All Types of Systematic Syntheses Applied to Meta-Reviews

Before continuing further into defining the scope of the meta-review, it is important to outline in brief the six major steps in conducting one. Meta-review authors should begin by developing and refining a protocol that attends to all planned actions for the six major steps. The protocol delineates the scope of the review and the steps that will be taken during the review process. To be trustworthy, the protocol must detail as many *a priori* decisions as possible, in order to reduce bias in conducting the review. The protocol should include the following components, all of which are unique, but interrelated steps in the review process: description of the problem of interest, search strategy (including electronic databases/hosts that will be searched and key terms, any websites or grey literature sources that will be searched and how), inclusion and exclusion criteria for the meta-review, the primary and secondary research questions, the process of literature screening and data extraction that will be undertaken, how dated-ness and overlap will be handled, what tool will be used to assess review quality, and the planned synthesis approach. Decisions for each step may vary depending on the scope of the meta-review as well as the team's resource capacity, size of the existing synthesis base, overlap among similar reviews, and degree of expected heterogeneity (Whitlock et al., 2008 in Cornell & Laine, 2008). As meta-reviews are a form of systematic review, including scholars who are experienced and skilled in systematic review methodology will save significant time, energy, and other resources.

Once the review protocol is developed, ideally, it should be registered or published in one of the available online repositories before the work commences. If the meta-review concerns a health intervention, registration on the website PROSPERO² is timestamped, free, publicly available, and the record can also be updated throughout the review process to mark milestones. Alternatively, if publishing a review with for example, the Cochrane, Campbell, or Joanna Briggs registries, the protocol will undergo a peer review process prior to registration, which may further refine the methods used. Finally, some journals will peer-review and publish protocols (e.g., BMJ Open, World Journal of Meta-analysis, and BMC Systematic Reviews). A meta-review is a collection of empirical work already publicly available, and thus we cannot overstate the important step of developing and pre-registering a protocol.

²<https://www.crd.york.ac.uk/prospero/>

Define Scope of the Meta-Review

A large part of developing the protocol involves formulating the research question and setting the meta-review scope. As we conclude elsewhere for systematic reviews in general (Johnson & Hennessy, 2019), setting an appropriate scope will ease the steps that follow. In setting the scope for a meta-review, four primary questions should guide the approach: (a) When is the literature a prime candidate for a meta-review? (b) What makes a good topic for a meta-review? (c) If they exist, what have past meta-reviews done well and what have they done poorly? What have meta-reviews ignored? And (d), what is existing team capacity to conduct a meta-review? The first three questions hinge on knowledge of the existing literature base and requires thoughtful attention to substantive content. One should also have a general understanding of the amount of both primary and synthesis studies in the area of interest as well as what questions these studies have answered.

A meta-review can encompass many different angles and scope. The more specific the meta-review question, the easier it will be to use conventional review guidelines; however, the broader the scope, the more amenable the meta-review will be to adaptations during the process, which may enable broader generalizations to different populations and contexts (Papageorgiou & Biondi-Zoccai, 2016). Meta-review authors have addressed a wide array of subjects: the use of methodological quality in health-related meta-analyses (e.g., Johnson, Low, & MacDonald, 2015); to examine why similar reviews reach disparate findings (e.g., Ebrahim, Bance, Athale, Malachowski, & Ioannidis, 2016); comparing and contrasting effects of behavioral interventions by pooling and re-analyzing trends from multiple meta-analyses (e.g., Johnson, Scott-Sheldon, & Carey, 2010); and, to explicate or develop theoretical frameworks (e.g., Protogerou & Johnson, 2014). There is not necessarily a “wrong” focus for a meta-review, but given the resources necessary to conduct a rigorous one, it seems that authors should focus on bridging literatures or illuminating discrepant findings. For example, if the meta-review examines whether health behavior interventions are effective, on average, for certain populations of individuals among reviews that focus on different population types, the review team would collect all types of health behavior intervention studies without distinction between populations, leaving the type of population as a key variable to code and analyze. From a methodological perspective, if the focus is on the history of psychology literature over time for studies involving, for example, the Theory of Planned Behavior, one could assess review and primary study quality over time and identify particular methods advances or address potential bias, for example, by exploring source of funding and review conclusions. Ideally, such meta-reviews could lead directly to plans for a new primary literature synthesis from the existing literature.

Although last on the list of questions, an accurate assessment of a team’s capacity for conducting a meta-review is vital to ensuring the project is feasible and encompasses several elements. First, the use of “team” is intentional: There should be multiple members to enable double-checking of steps taken throughout the process (Higgins & Green, 2011; Johnson & Hennessy, 2019). Members on the review team must be comfortable with synthesis literature and the way it is reported, and if the protocol includes the synthesis of meta-analyses, then at least one author who has conducted a meta-analysis should be included so they can address the range of quantitative issues raised in meta-analytic research: Accurately identifying areas

of strengths and weaknesses in the analytic approach and how these issues contribute to implications of the results is vital to an accurate assessment of the trustworthiness of the results. Although some standardized tools can assist this process (covered in the section, **Assessing Review Quality**), an involved team member should be versed in these issues to identify nuances that the tools may not easily identify but that could affect the findings. At least one member of the team should be someone who is comfortable with the substantive topic area as well as those who have a good understanding of the scope of the literature and of the best places to identify relevant literature. This teamwork will help in defining the areas for research questions, operationalizing concepts for a protocol and data extraction, identifying relevant sources of literature, clarifying the search terms for search strategies, and reporting findings in a way that is most useful for the particular focus of the meta-review.

Time and software are also important resource needs to consider. From a sample of systematic reviews, Allen and Olkin's (1999) model predicted 721 hours of start-up time (protocol development, searches, and database development). Using their equation, one can estimate the time needed to conduct a review after identifying potentially eligible citations: For example, a search that returned 2,500 citations should involve approximately 1251 hours of review task time. Similarly, Borah and colleagues (2017) estimated that for reviews of interventions, if reviews range from 27 to 92,020 studies, the average rate of completion would be over a year ($M = 67.3$ weeks, range = 6 to 186 weeks). Furthermore, given the nature of many review tasks, the overall timeframe for completing a review often extends beyond a year because those undertaking reviews typically have multiple competing tasks and deadlines, especially on teams composed of both practitioners and researchers. Although these are estimates for conducting a primary evidence synthesis, given the overlap in tasks, it seems likely these estimates translate well to approximating the time needed to conduct a meta-review. As unfinished reviews are a waste of valuable resources, we encourage the use of online tools to help with accurate planning (e.g., <https://estech.shinyapps.io/predicter/>). Savvy use of the right software can also save time in many of the meta-review tasks (e.g., managing potentially relevant reviews and their references, identifying and handling reference duplicates, allowing for multiple rounds of double and independent screening and data extraction, and by enabling a standardized data extraction form that is both easy for data collection and outputs results in a format easy to synthesize and use in the final report). Some programs are free but cumbersome to use, while others are expensive and may require experts to design and implement; these are important considerations about a team's capacity to conduct a review that should be considered during the setting the scope process.

Ensuring Literature Sources and Data Collected Align with the Scope

In setting the meta-review scope, it is important to attend to inclusion and exclusion criteria as one would in a primary literature synthesis. Often as the review process proceeds, authors may revisit the size of the scope and ensure the criteria match the scope or make the appropriate changes to reset the scope. Setting inclusion and exclusion criteria for meta-reviews parallels those used for systematic reviews or meta-analyses, using for example, CIMO (Denyer, Tranfield, & Van Aken, 2008), SPIDER (Cooke, Smith, & Booth, 2012), or as we recently recommended, TOPICS+M (Johnson & Hennessy, 2019), which includes

potential moderating factors of interest in addition to the traditional elements of population, intervention/treatment, comparator, outcomes, and study design elements; see these citations for additional details on using these frameworks.

Meta-review authors must also consider whether all types of syntheses will be eligible for the review, or if there will be a restriction to a certain type of analysis (e.g., meta-analyses) and to justify why the restriction is in place. For example, a meta-review focused on how review authors statistically examined publication bias would only include quantitative syntheses. Similarly, a meta-review of the effectiveness of psychological interventions to address health outcomes would also include only quantitative syntheses, as narrative syntheses would be unable to generate findings on effectiveness aside from vote-counting, which is not an optimally rigorous approach to synthesis. Alternatively, a meta-review focused on the experience of individuals in a particular treatment may best be approached from a meta-synthesis of qualitative reviews or by including both qualitative and quantitative reviews to address experiences and effectiveness.

Other important considerations for setting the meta-review inclusion criteria include: (a) whether there will be limits placed on dates of included reviews or their inclusion of studies; and (b) to ensure up-to-datedness of the literature, although this concern may be primarily addressed with the search filter. In some areas with many meta-analyses, limiting to recent ones is indicated given that methodological quality keeps improving, although attention should be paid to the comprehensiveness of reviews. Also, whether any criteria set by the reviews, e.g., the type of study design (randomized controlled trial, quasi-experimental), participant population, or setting, can be considered for the meta-review inclusion criteria. Finally, it is possible to use the quality of the included reviews as a secondary exclusion criterion, making it an example of *best-evidence meta-review*; accordingly, reviews deemed eligible after the first round of screening are subjected to the quality assessment tool and only those meeting a specific *a priori* quality threshold are included. An advantage to this approach is that there is some assurance of high quality of evidence in the final meta-review. Potential disadvantages of the approach are (a) use of arbitrary criteria for inclusion; (b) not being able to determine whether reviews with lower quality reach the same conclusions as reviews with higher quality; and (c) if the scope of the meta-review is to map the state of an entire literature, then best-evidence meta-reviews will likely leave important gaps. (We provide more detail on quality assessment below; see **Assessing Review Quality**). Of note, expanding the scope increases the work and resources involved in completing the review, so it is important to ensure the review scope takes those limitations into account.

Once the scope is set, there are three steps that should be involved in all systematic research syntheses: (1) engaging in a comprehensive and systematic search process, (2) conducting careful screening and data extraction that minimizes error or potential for bias, and (3) appropriately assessing potential of risk of bias or methodological quality. As with setting the scope, these steps are quite similar across different types of systematic research syntheses so will not be reviewed in detail here; instead, this section will briefly address what we think are the most important points and then focus on particular adaptations that are useful to consider when conducting a meta-review.

Comprehensive and systematic search process.—As with all systematic reviews, a meta-review must have a systematic, transparent, and reproducible search (Biondi-Zoccai, 2016; The Cochrane Collaboration, 2011). Moreover, a search that is comprehensive and well implemented will maximize the identification of relevant syntheses available through searching varied and multiple sources. Both the Campbell Collaboration and the Cochrane Collaboration have published extensive guidelines for information retrieval in systematic reviews (Kugley et al., 2016) and these advocate consultations with an information retrieval specialist from the outset. Such experts understand the complexities associated with uncovering relevant literature through systematic searching and are equipped to assist with developing the search strategy, including choosing the best databases and determining their relevance to the research question. Information retrieval specialists can also help to identify the key concepts of the meta-review that can be turned into precise search terms.

A common theme across all meta-reviews is that the study design to be included will be an evidence synthesis product (systematic review or meta-analysis). It is therefore imperative that searches cover those databases that house evidence synthesis products exclusively, including any relevant to the topic of the meta-review, such as the Campbell Collaboration's library, the Cochrane Database of Systematic Reviews (CDSR), and the Database of Abstracts of Reviews of Effects (DARE). However, relevant syntheses may also be unique to other databases such as PubMed, Embase, and PsycINFO; therefore, it is advisable to utilize as many relevant subject databases as possible, not only to ensure that relevant reports have been located, but also to reduce potential selection bias (Dickersin, Scherer, & Lefebvre, 1994).

The 'pearl harvesting' (PH) method has proven effective in locating the most relevant and inclusive keywords across various key concepts (Sandieson, Kirkpatrick, Sandieson, & Zimmerman, 2010). The method follows exact guidelines to find all the necessary keywords to locate relevant syntheses. When searching for relevant syntheses for meta-reviews, the key concept of study design is represented by as many terms as are necessary to capture all the relevant articles, including those with titles where a synthesis term was misspelled (e.g., "metaanalysis" with no hyphen). Therefore, by copying and pasting the comprehensive list of methodological terms presented in the online supplement (Figure S1) alongside other key concepts of interest (e.g. population and/or intervention), meta-review teams can quickly uncover and collate relevant reports and avoid sorting through extensive primary studies which would be ineligible for a meta-review. The terms in this Figure have been refined and implemented in both the Proquest and OVID interfaces and therefore can be used as a starting point in searches of such databases as PsycINFO, Dissertation and Theses Global, ERIC, MEDLINE, EMBASE and others.

As a result of setting the meta-review's scope, some common limits that are applied in the search process relate to year of publication (e.g., those published after a certain date) and language of report (e.g., those published in English). When choosing a date range for possible inclusions, authors should be careful to not choose a period that is so narrow that eligible studies are limited and therefore interpretation is restricted. Similarly, they should not choose a period so wide that the meta-review is not feasible or not representative of current practice. Meta-review teams should not exclude reports written in another language

as it may increase bias and reduce precision of findings through the exclusion of useful and relevant data (Neimann Rasmussen & Montgomery, 2018). Online translation software has become increasingly quick and accurate and can put text into the team's preferred language (with the caution that the software may miss nuance due to the original text's inclusion of idiomatic language). If meta-review teams still decide to exclude reviews outside of their primary language, then potentially eligible reviews from other languages that were identified during the screening process should be listed so that future meta-reviews have them as a starting point for their own samples. Meta-review authors should report and justify any filtering decisions that are not substantively related to the question of interest (Meline, 2006).

Screening and data extraction.—Guidelines for traditional systematic reviews should be followed when engaging in screening and data extraction processes. That is, to ensure as few errors as possible are introduced during screening, it is best to double-screen independently and in duplicate at the title/abstract level first, and then repeat the process with full-text reports of potentially eligible studies. Of course, if resources are insufficient for complete duplication of effort, then some strategies can save time. For example, during the title/abstract screening, one way to speed up this process is to include any item as potentially eligible as long as one person has included it, rather than conduct discrepancy resolution. Yet, if the search has returned a large number of citations, it may make the next steps of the process (retrieval and full-text screening) much longer. One could also use a liberal screening process whereby one team member reviews all reports and a second team member only reviews items excluded by the first person; this process ensures that the search does not exclude any report that should be included. When conducting screening for a meta-review, the team should ensure there is a way for screeners to tag relevant meta-reviews so they will be collected and their bibliographies reviewed for potentially eligible studies. Finally, we envision a future in which the “drudge” of literature search and report selection will be significantly reduced via the use of machine-learning strategies or even crowd sourcing (Marshall, Noel-Storr, Kuiper, Thomas, & Wallace, 2018; Martin, Surian, Bashir, Bourgeois, & Dunn, 2019; Mortensen, Adam, Trikalinos, Kraska, & Wallace, 2017).

Data extraction should always be guided by a standardized form and one that the team has piloted before fully implementing it. The meta-review team should create a codebook to define key distinctions; then coders can use these definitions to ensure accurate coding, updating the codebook as needed when confusing instances arise. As with screening, it is best to double-code independently and in duplicate and resolve discrepancies as they arise. Meeting often throughout this part of the review process will ensure that the team catches data extraction errors early. Last, it is important to consider how to handle missing data and whether it is a result of faulty reporting, different scope, or different disciplinary traditions. Decisions for how to address missing data may take different steps, depending on the type of missing data. For example, in some cases it will be necessary to email authors, whereas in others the missing information may be available in linked study reports (e.g., if a review protocol was registered there may be additional methodological-related information located there). In conducting our own meta-reviews, we have identified several distinct elements to

consider as part of a data extraction sheet for meta-reviews as well as potential ways the answer choices could be structured, and provide these suggestions in Table 1.

Assessing Review Quality

Quality of the underlying literature in a meta-review influences all implications from the findings and includes attention to issues in primary studies, the reviews that synthesize them, and ultimately the finished meta-review. Although authors have a number of checklists to help them meet reporting and quality standards for reviews, including Methodological Expectations of Cochrane Intervention Review (Churchill, Lasserson, Chandler, Tovey, & Higgins 2016), US Institute of Medicine standards for high quality systematic reviews (Institute of Medicine, 2011), or the PRISMA statement and checklist (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009; Moher et al., 2015), none of these are formalized quality assessment tools for individuals using those reviews. Additionally, reporting quality—the level of detail of the synthesis report—will determine how reliable internal and/or external validity assumptions are for a particular review. Given these issues, in a meta-review it is important to consider not only the quality of the primary studies included in the included reviews but also the methodological limitations of the reviews themselves in order to determine the overall quality of the existing evidence.

To begin with a quick assessment, some items that are easy to note are (a) whether the authors complied with a reporting standard such as PRISMA and included it with their publication; (b) whether the review protocol was registered before the review commenced; (c) whether the protocol was detailed; and (d), whether the authors followed the protocol and described any deviations from it. Beyond these preliminary steps, a formalized assessment of review quality should occur using one of the established tools for this purpose. As there are currently over 20 tools available, there seems little reason for meta-review authors to create their own. Instead, teams would do well to choose one from inventories designed and validated to assess methodological quality of systematic reviews. We name two that are widely used, especially for research using experimental designs: There are the two AMSTAR inventories (AMSTAR 2: Shea et al., 2017; AMSTAR: Shea et al., 2007) and Risk of Bias in Systematic Reviews (ROBIS: Whiting et al., 2016). These inventories address a number of quality issues and potential areas for risk of bias and include questions to examine whether reviews properly assessed the research methods of included primary studies. Both appear useful for reviews that focus on research that does not use experimental designs, although some questions may need to be adapted to suit any other research design assumptions; alternatively, authors may wish to supplement these scales with the other validated inventories (e.g., (Critical Appraisal Skills Programme, CASP, 2019). These instruments should be carefully applied and their use described in detail by meta-review authors (Pieper, Koensgen, Breuing, Ge, & Wegewitz, 2018), including whether modifications were applied. Similar to the process for data extraction, meta-review teams should engage in training of assessors and a pilot implementation period of these tools to ensure their application is accurate.

After using a standardized assessment, authors must consider the most appropriate way of relaying trends. Although authors may choose to create a summary score in the interest of

simplified reporting, these scales were not intended to be used in this way and simply presenting a scale score without discussing the major issues and implications of these findings oversimplifies the issue (Cornell & Laine, 2008; Johnson, et al., 2015; Valentine & Cooper, 2008). Because rating scales may address *reporting quality* rather than *methods*, authors should acknowledge this potential limitation, or, if resources permit, contact review authors to ask if they used particular methods. Because of these potential limitations and likelihood for subjectivity, a meta-review team should ensure that quality is assessed independently and in duplicate, with discussions to resolve any discrepancies, and that reporting is transparent. Appendices should be used as necessary to ensure all relevant items and rationale for the ratings are reported.

Meta-review authors must attend to the potential for reporting bias, which becomes compounded across study levels (McKenzie, 2011; Page, McKenzie, & Forbes, 2013). At the primary study level, for example, multiple outcomes can be measured/analyzed but only a subset are reported (e.g., if only some are significant) and then at the review level, only a subset of outcomes may be reported (and/or multiple outcomes are analyzed, but only a subset are reported). This issue can initially be addressed through assessing whether review authors looked for potential reporting bias among included primary studies. Second, by examining whether the authors registered a protocol that pre-specified the primary and secondary outcomes of interest and if so, whether they deviated from it can be informative. Given that meta-reviews of both the Cochrane Library and the PROSPERO database have identified a substantial portion of reviews that had not pre-specified a primary outcome or, alternatively, had changed outcome reporting after conducting the review (Kirkham, Altman, & Williamson, 2010; Tricco et al., 2016), this practice is vital to ensure potential for future replication in any scientific field. Finally, it is important to realize that a meta-review itself can have high quality methods even if the reviews it synthesizes lack it; further, such a review is likely not to be a synthesis of reliable results but instead to be a statement about state of scientific methods in the particular research domain.

Meta-Review Issues and Methods to Address Them

Datedness

The datedness of the literature in the meta-review is important to consider, because as Pattanittum and colleagues argued, “Evidence is dynamic, and if systematic reviews are out-of-date this information may not only be unhelpful, it may be harmful” (Pattanittum, Laopaiboon, Moher, Lumbiganon, & Ngamjarus, 2012, p. 1). With meta-reviews specifically, datedness is important at multiple levels. First, consider the search dates for the included reviews and the dates of the primary studies synthesized in those primary reviews. That is, a systematic review could have comprehensively searched the older literature so may include interventions that are no longer used in clinical practice (e.g., Shojania et al., 2007). Meta-review authors can prevent out-of-datedness by using search or inclusion criteria based on knowledge of the state of the literature and by updating reviews every few years (Cooper & Koenka, 2012). Alternatively, some quantitative methods can assist with this task (e.g., see Pattanittum et al., 2012). Meta-review teams can also scrape the dates of included studies from extant meta-analyses to produce a temporal model to predict how many studies may

have appeared following the most recent review's search. For example, if a linear model is reasonable, a review team might find that:

$$\hat{k} = -1086.142 + 0.5531136 \times \text{Year},$$

where \hat{k} is the predicted number of studies in a particular year. The team would then apply the equation to all relevant years, from the first year research on the subject appeared until the current year, and the sum of these values predicts how many studies likely exist. Such a calculation might show, for example, the most recent review is likely to be decidedly out of date. Obviously, the meta-review team will have to evaluate whether a linear model makes the most sense; some literatures might be characterized by accelerating or decelerating curves over time. Whichever method is chosen to address datedness of the literature, meta-review authors should acknowledge the likelihood that some of the included literature may be out of date, and they should be proactive in discussing key advances in the field that readers should consider, for example by drawing from relevant, recent primary study literature (Polanin, Maynard, & Dell, 2017).

Importantly, if there have not been recent systematic reviews in the area of interest, then the existing systematic reviews may not be current enough to synthesize; in this case, conducting a new systematic review that updates the searches from previous reviews, rather than a meta-review is more appropriate. In such a case, the study team should consider whether it has the substantive expertise necessary to conduct this review and then to decide how it should be updated to improve on existing reviews. A scoping search (Cooper & Koenka, 2012) might also help to identify newer studies that have innovated new methods and improve the eventual systematic review focused on the phenomenon.

Overlap

Overlap in a meta-review occurs to the extent that two or more syntheses cover the same primary studies (Pieper, Antoine, Mathes, Neugebauer, & Eikermann, 2014). Overlap occurs with reviews that are updated frequently, as the authors will often add new studies in addition to the original studies, or with reviews that cover similar topics but may have a different focus (e.g., emphasis on different moderators, expanded inclusion criteria, alternative analysis methods). Overlap, if ignored by meta-review authors is problematic: Just as meta-analyses assume non-independence between outcomes (or statistically control for it), one must not to overestimate the importance of an effect in a meta-review by including review findings from the same primary studies multiple times.

Pieper and colleagues (2014) provide an easily implemented solution to describe the overlap in a meta-review – calculate the *corrected covered area* (CCA). First, create a citation matrix with primary publication citations (one per row) and individual reviews (one per column); then, after sorting alphabetically and removing duplicate publications, mark a check for each primary publication that is included in each review. Then,

$$CCA = \frac{k - r}{r(c - r)},$$

where k is the number of reports in reviews (sum of ticked boxes), r is the number of rows (index publications), and c is the number of columns (included reviews). Pieper and colleagues (2014) suggest criteria for interpreting the index, with slight (0–5), moderate (6–10), high (11–15) or very high (>15) overlap. In our own work, we have found it easiest to create an initial matrix in a spreadsheet to map references across reviews and provide counts, although it seems likely future reference management or database software will automate and circumvent this tedious task.

Depending on the focus of the meta-review, authors may find it necessary to calculate the CCA both across all reports included and for particular outcomes (Pieper et al., 2014). That is, if there are seemingly discrepant findings between reviews, then meta-review authors could examine overlap for reports addressing the same outcomes, to identify why discrepancies are present. For example, in our recent meta-review of self-regulation mechanisms of health behavior change interventions (Hennessy et al., 2018), we found some evidence that personalized feedback was an effective intervention component for improving physical activity, but this mechanism was inconsistent across three meta-analyses. By examining the CCA across the three reviews (<1%) and the shared reference lists, as well as the aims and populations, we reached a conclusion: The discrepancy appears to have occurred due to differences in included populations (one focused on older adults, a second had no population restrictions, and a third focused on overweight and obese adults). In other instances, the discrepancy could be due to type of effect sizes used, methods of outcome synthesis, or some other review team decision; meta-review teams must address these areas when integrating the literature.

Aside from calculating the CCA, it also may be useful to simplify the presentation of findings by removing some studies or some syntheses. To explain, if syntheses overlap to a high degree, then the one with the most information or largest number of studies could be prioritized over the others; alternatively, newer or higher quality syntheses could be prioritized (Pollock, Campbell, Brunton, Hunt, & Estcourt, 2017). The exact methods used to address overlap will vary by review scope and the particular outcomes of interest: Meta-review authors should always detail methods chosen to address overlap, ideally at the protocol stage but certainly in the final manuscript. Additionally, if reviews were eligible according to inclusion criteria but were removed due to overlap, meta-review authors should report these citations.

Synthesis Options

Prior to presenting outcome findings, meta-reviews should generally include descriptive information about each of the included reviews, as well as information on review quality. When it comes to reporting key outcomes, authors of meta-reviews have three primary synthesis options (Papageorgiou & Biondi-Zoccai, 2016): (a) narrative synthesis; (b) “semi-quantitative” synthesis”; or (c) quantitative synthesis. Of course, the choice of synthesis option will vary by the aim of the review scope and the type of literature reviewed; it will also depend on the team’s familiarity with quantitative syntheses, the type of assumptions the team is willing to make about the underlying data, and of course, the quality and heterogeneity of the primary studies in each review. Ideally, meta-review authors would

decide before beginning the meta-review what type of synthesis to conduct, but this choice should be revisited if the literature is vastly different than what was expected.

Narrative synthesis.—The narrative synthesis approach may seem intuitive to many meta-review authors, yet the analysis and reporting must be conducted systematically. We briefly address one common method that could be used with reviews of diverse natures, including those with quantitative information³: content analysis of the text, which is a systematic way of analyzing text to “attain a condensed and broad description of the phenomenon, and the outcome of the analysis is concepts or categories describing the phenomenon” (Elo & Kyngas, 2008, p. 108). This approach involves three main phases:

1. In the *preparation phase*, reviewers select an appropriate and representative unit of analysis from the ‘universe’ from which it is drawn. Preparation next requires data immersion by reading the reviews in depth multiple times to “mak[e]sense of the data”, i.e., gain a clear understanding of ‘who’ is reporting; ‘where’ it is happening; ‘when it happened; ‘what’ happened; and ‘why’ (Burnard, 1996; Dey, 1997; Duncan, 1989; Elo & Kyngas, 2008).
2. Next, in the *organization phase*, open coding, creating categories, and abstraction occurs. Open coding entails writing notes and headings in the margins of the review articles, while reading them across several iterations, to describe as many aspects of the reviews as possible. During this stage, authors should look for both convergence and divergence between different methods used to synthesize primary study literature.
3. Finally, in *abstraction*, authors will more systematically formulate a general description of the data by generating categories and grouping features together by giving them an appropriate label, which captures the essence of those features. This synthesis method is described as a cyclical process and may continue as far as reasonable and as far as possible. Table 2a shows an example of this process.

Semiquantitative synthesis.—A semiquantitative synthesis involves converting summary statistics to a common metric but does not conduct a formal meta-meta-analysis of the data. For example, some meta-reviews interested in understanding publication bias in a field have estimated *z*-scores to indicate the difference between published and unpublished effect sizes, to examine factors linked to bias. Alternatively, meta-reviews have focused on examining correlations between review quality or other characteristics and reported outcomes. Table 2b provides some examples of meta-reviews and a variety of these semi-quantitative methods.

Quantitative synthesis.—Quantitative synthesis options for meta-reviews are not yet well-developed, although Schmidt and Oh (Schmidt & Oh, 2013) distinguish between three

³Although we briefly cover a common way to narratively synthesize findings, many qualitative literature and qualitative syntheses may require a more refined approach; other sources provide more in-depth coverage of these methods (For a more in-depth critical review of these other methods see Barnett - Page & Thomas, 2009; Thorne, Jensen, Kearney, Noblit, & Sandelowski, 2004).

other types of meta-meta-analyses: Only those well-versed in meta-analytic assumptions should undertake the approaches that follow.

1. An *omnibus meta-analysis* involves pooling all primary studies across all moderators, and then conducting separate meta-analyses for moderators. This is only possible if all primary studies and data used in each first order meta-analysis are available. Thus, this option will not be feasible in all disciplines, can be time consuming, and the variance estimate (and percentage of that variance) due to second order sampling error cannot be estimated.
2. The second method is the most common; that of averaging mean effect sizes across first order meta-analyses. This method ignores between-meta-analysis variance and assumes that the included meta-analyses have minimal overlap. Although meta-analyses routinely report such quantitative statistics, limitations with this approach include problems such as an inability to estimate the “true” variance between meta-analytic means, credibility intervals for second-order meta-analytic means, amount of observed variation across meta-analyses due to second-order sampling error, or confidence intervals for second-order meta-analytic means.
3. Schmidt and Oh (2013) advocate for an approach that combines mean effect sizes across meta-analyses and models between-meta-analysis variance. This approach assumes that there are random effects within and across meta-analyses, inverse variance sampling weights, and sample independence. However, tools for conducting this method of synthesis are not fully developed or perfected and there is no fixed effect equivalent because FE assumes no “real variation” in effect sizes across studies beyond sampling error.
4. We add a final fourth approach, which is combining individual studies’ effect sizes from all meta-analyses in the sample (but only incorporate them once), and then model the effects using either multi-level meta-analysis or robust variance estimators, which both control for clustering of non-independent observations, if applied correctly (Hedges, Tipton, & Johnson, 2010; Moeyaert et al., 2017; Tanner-Smith & Tipton, 2014).

Considerations for effect size data.—Effect size data are the “coin of the meta-analytic realm” (Rosenthal, 1995, p. 185), a sentiment highly relevant for meta-reviews incorporating quantitative findings. Yet, evidence indicates that even meta-analyses are frequently opaque with regard to quantitative information and can present conflicting, incorrect, or inadequate data (Cooper & Koenka, 2012; Polanin, Hennessy, & Tsuji, 2019; Schmidt & Oh, 2013). For reviews that present all the effect size data necessary for replication, meta-review authors must assess whether the review is trustworthy enough to take data at face value (e.g., through use of a review quality tool or via random spot check of presented effect sizes). When effect sizes appear incorrect, it may be necessary to contact authors of the review because the alternative approach would be to recalculate effect sizes, a considerable burden. Increasingly, efforts are underway to make meta-analyses more transparent and reproducible; once again, computer-based machine learning appears to offer

hope in reduced effort in extracting data from reports (Michie et al., 2017). Thus, we are hopeful this issue will eventually become a footnote of history.

Reporting Meta-Review Findings

Reporting synthesis results is fairly straightforward for a meta-review, and authors of primary systematic reviews should be familiar with the process. There are a few key considerations that meta-review authors should address. First, meta-reviews should also be easily identifiable, by including this term or a discipline-appropriate synonym and further recommendations include that as many of the PICOS or TOPICS+M elements in the title as possible are included (Papageorgiou & Biondi-Zoccai, 2016).

Second, meta-reviews should identify area(s) for future high quality systematic reviews or meta-analyses, or, if they found a number of these already on the topic, then a potential primary study driven by emerging methodological approaches or alternative moderator specifications (or other further under researched questions of interest). In this task, meta-review teams must also be careful not to overreach: Any implications for practice should be practical, unambiguous, and justifiable by data presented in the review (Becker & Oxman, 2008). Meta-review teams should also carefully describe results and use the more accurate phrase 'No evidence of effect' instead of 'Evidence of no effect'. Furthermore, a few other key elements, if included in the meta-review, help ensure a comprehensive report: (a) Characteristics of included reviews table (see example Overview of reviews table in Becker & Oxman, 2008); (b) a flow diagram to detail the search process; (c) review quality or risk of bias assessment results (aggregate figure of overall items and a table with ratings for each individual study on each item); (d) a reference list of all included reviews, and (e) outcome data.

Depending on the outcome(s) of interest, there are a few different ways to present results. Authors have suggested a harvest plot-matrix style plot to visualize multiple variables at once (Ogilvie et al., 2008) or a modified Forest plot/Forest top plot (Becker & Oxman, 2008) to engage readers and offer a concise visual summary. Those figures, especially the Forest top plot work well if the outcomes are on a similar metric. As we demonstrate in in the supplemental files (Figure 2), scatterplots are another way to represent a summary of findings with relevant detail. For example, this scatterplot drawn from a recent meta-review of mechanisms of health behavior change interventions, provides information on five key characteristics synthesized in the meta-review including (1) methodological quality information (X-axis) and (2) supportiveness of meta-analyses supplied in (3) favor (plot a) or opposed (plot b) for (4) individual self-regulation mechanisms (listed on Y-axis) for all reviews. Bubbles for each meta-analysis are sized proportional to the (5) numbers of studies each analysis included and their colors indicate the focal health behavior of each analysis. When viewed in this format, readers can quickly gauge the size and quality of the synthesis literature base for certain intervention techniques and for certain health behaviors.

Future Directions: Advances in Meta-Review Synthesis Methods

Although synthesizing literature is not a new phenomenon, there are a number of areas for future development in meta-review methodology to meet the demands of an increasingly

complex literature base. For example, although authors could use existing guidelines for primary systematic reviews, a systematic quality rating of meta-reviews tool must be established. Such a tool would attempt to rate how well the authors handled issues specific to meta-reviews, including up-to-datedness of the review literature, overlap between reviews, and synthesis approaches. Additionally, the GRADE tool is currently being adapted for use in meta-reviews; thus, readers will be encouraged to use that system, as it is refined and becomes available (Schünemann, Broek, Guyatt, & Oxman, 2013). This essay has only briefly addressed quantitative synthesis options available for meta-reviews, as these methods are still being developed. The last 20 years have seen a marked increase in the statistical strategies for primary research syntheses and it is only a matter of time before options for meta-reviews develop as well.

Synthesis authors increasingly use complex methods for combining primary study results, such as robust variance estimation, multilevel modelling, network meta-analysis, and structural equation modelling, among others. Their existence begs the question of how meta-review authors should examine and synthesize these. Given the rise in environmental-level data, authors should also begin to think about their research questions in more spatiotemporal terms. All issues of research interest occur in space and time; some of these variables may explain more variance than some of the traditional moderators – or in interaction with these moderators. Thus, methods to incorporate spatiotemporal considerations into research syntheses are needed to address the true complexity of study findings in psychology and other fields (Johnson, Cromley, & Marrouch, 2017).

Finally, transparency of original reviews must improve; authors of such reviews should be encouraged to pre-register their protocol and document any *post hoc* changes that occur, consistent with the feedback loop in Figure 1. They also should include effect size data from individual studies, if the journal allows it (e.g., in online supplements or in separate archives).

Conclusions

This article aimed to lay out the steps and options for the best conduct of a systematic meta-review in applied psychology. Here we have described the potential roles of meta-reviews and the primary reasons why they should be conducted. We have presented appropriate options for addressing meta-review challenges, such as conducting thorough searches and incorporating rigorous methodological quality assessments of the included reviews. Indeed, although important to pre-specify as much as possible *a priori*, similar to a primary study, if things are not going according to plan, modifications should be made – and transparently noted – for readers. And, we have outlined areas for advances in meta-review synthesis methods. The details we have provided point readers in the correct direction, but while these steps represent a systematic way of conducting a meta-review, authors should not blindly follow these as rules. Despite the minutiae and nuance necessary to conduct a high-quality meta-review, because they can organize a morass of crucial details, we believe that meta-reviews will offer increasingly influential statements about the state of the science across scientific fields, guiding the conduct not only of new systematic reviews but also of new original studies in these domains. Our hope is that this article has given readers a sense of

the breadth of options available and the background knowledge to make the best decisions for teams interested in conducting meta-reviews and those consuming them.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Details:

This work is supported by the National Institutes of Health (NIH) Science of Behavior Change Common Fund Program through an award administered by the National Institute on Aging (U.S. PHS grant 5U24AG052175). The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

- Allen IE, & Olkin I (1999). Estimating time to conduct a meta-analysis from number of citations retrieved. *JAMA*, 282(7), 634–635. doi:jlt0818 [pii] [PubMed: 10517715]
- Barnett - Page E, & Thomas J (2009). Methods for the synthesis of qualitative research: A critical review. *BMC Medical Research Methodology*, 9(1), 59. doi:10.1186/1471-2288-9-59 [PubMed: 19671152]
- Bastian H, Glasziou P, & Chalmers I (2010). Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? *PLoS Medicine*, 7(9), e1000326. doi:10.1371/journal.pmed.1000326 [PubMed: 20877712]
- Becker LA, & Oxman AD (2008). Overviews of reviews. In Higgins JP, & Green S (Eds.), *Cochrane handbook for systematic reviews of interventions: Cochrane book series* (pp. 607–631) Chichester, England: Wiley-Blackwell.
- Biondi-Zoccai G (Ed.). (2016). *Umbrella reviews: Evidence synthesis with meta-reviews of reviews and meta-epidemiologic studies* (1st ed.). Switzerland: Springer International Publishing.
- Borah R, Brown AW, Capers PL, & Kaiser KA (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, 7(2), e012545–2016-012545. doi:10.1136/bmjopen-2016-012545
- Borenstein M, Hedges LV, Higgins JP, & Rothstein HR (2011). *Introduction to meta-analysis* Chichester, UK: John Wiley & Sons.
- Burnard P (1996). Teaching the analysis of textual data: An experiential approach. *Nurse Education Today*, 16(4), 278–281. doi:10.1016/S0260-6917(96)80115-8 [PubMed: 8936234]
- Churchill R, Lasserson T, Chandler J, Tovey D, & Higgins JPT (2016). Standards for the conduct of new Cochrane Intervention Reviews. In Higgins JPT, Lasserson, Chandler J, Tovey D, & Churchill R (Eds.), *Methodological Expectations of Cochrane Intervention Reviews* London, England: Cochrane.
- Cooke A, Smith D, & Booth A (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, 22(10), 1435–1443. [PubMed: 22829486]
- Cooper H, & Koenka AC (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *The American Psychologist*, 67(6), 446–462. doi:10.1037/a0027119 [PubMed: 22352742]
- Cornell JE, & Laine C (2008). The science and art of deduction: Complex systematic overviews. *Annals of Internal Medicine*, 148(10), 786–788. doi:148/10/786 [PubMed: 18490692]
- Critical Appraisal Skills Programme, CASP. (2019). *Qualitative research checklist* Retrieved from https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf
- Denyer D, Tranfield D, & Van Aken JE (2008). Developing design propositions through research synthesis. *Organization Studies*, 29(3), 393–413. doi:10.1177/0170840607088020

- Dey EL (1997). Working with low survey response rates: The efficacy of weighting adjustments. *Research in Higher Education*, 38(2), 215–227. 10.1023/A:1024985704202
- Dickersin K, Scherer R, & Lefebvre C (1994). Identifying relevant studies for systematic reviews. *British Medical Journal*, 309, 1286–1291. doi: 10.1136/bmj.309.6964.1286 [PubMed: 7718048]
- Duncan DF (1989). Content analysis in health education research: An introduction to purposes and methods. *Health Education*, 20(7), 27–31. 10.1080/00970050.1989.10610182 [PubMed: 2516070]
- Ebrahim S, Bance S, Athale A, Malachowski C, & Ioannidis JP (2016). Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. *Journal of Clinical Epidemiology*, 70, 155–163. doi:10.1016/j.jclinepi.2015.08.021 [PubMed: 26399904]
- Elo S, & Kyngas H (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. doi:10.1111/j.1365-2648.2007.04569.x [PubMed: 18352969]
- Hedges LV, Tipton E, & Johnson MC (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. doi:10.1002/jrsm.5 [PubMed: 26056092]
- Hennessy EA, Johnson BT, Acabchuk RL, McCloskey K, Iqbal S, Roethke L, ... McDavid K (2018). Self-regulation as a target mechanism of behaviour change interventions: A meta-review. Paper presented at the Presentation at the 32nd Annual Conference of the European Health Psychology Society. Galway, Ireland.
- Higgins JP, & Green S (2011). *Cochrane handbook for systematic reviews of interventions* (5.1.0, Updated March 2011 ed.) The Cochrane Collaboration Retrieved from www.cochrane-handbook.org
- Institute of Medicine. 2011 *Finding What Works in Health Care: Standards for Systematic Reviews* Washington, DC: The National Academies Press 10.17226/13059
- Ioannidis JP (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, 94(3), 485–514. doi: 10.1111/1468-0009.12210 [PubMed: 27620683]
- Johnson BT, & Hennessy EA (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. Submitted for Publication
- Johnson BT, Low RE, & MacDonald HV (2015). Panning for the gold in health research: Incorporating studies' methodological quality in meta-analysis. *Psychology & Health*, 30(1), 135–152. doi:10.1080/08870446.2014.953533 [PubMed: 25137094]
- Johnson BT, Cromley EK, & Marrouch N (2017). Spatiotemporal meta-analysis: Reviewing health psychology phenomena over space and time. *Health Psychology Review*, 11(3), 280–291. doi: 10.1080/17437199.2017.1343679 [PubMed: 28625102]
- Johnson BT, Scott-Sheldon LA, & Carey MP (2010). Meta-synthesis of health behavior change meta-analyses. *American Journal of Public Health*, 100(11), 2193–2198. doi:10.2105/AJPH.2008.155200 [PubMed: 20167901]
- Kirkham JJ, Altman DG, & Williamson PR (2010). Bias due to changes in specified outcomes during the systematic review process. *PloS One*, 5(3), e9810. doi:10.1371/journal.pone.0009810 [PubMed: 20339557]
- Kugley S, Wade A, Thomas J, Mahood Q, Jørgensen AMK, Hammerstrøm K, ... Sathe N (2016). *Searching for studies: A guide to information retrieval for campbell systematic reviews* Oslo, Norway: The Campbell Collaboration. doi:10.4073/cmg.2016.1
- Lipsey M, & Wilson DB (2001). *Practical meta-analysis* Thousand Oaks CA, USA: Sage Publications.
- Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, & Wallace BC (2018). Machine learning for identifying randomized controlled trials: An evaluation and practitioner's guide. *Research Synthesis Methods*, 9(4), 602–614. doi: 10.1002/jrsm.1287 [PubMed: 29314757]
- Martin P, Surian D, Bashir R, Bourgeois FT, & Dunn AG (2019). Trial2rev: Combining machine learning and crowd-sourcing to create a shared space for updating systematic reviews. *JAMIA Open*, 2(1), 15–22. 10.1093/jamiaopen/ooy062
- McKenzie. (2011). *Methodological issues in meta-analysis of randomized controlled trials with continuous outcomes* (Unpublished PhD). Monash University, Melbourne, Australia.

- Meline T (2006). Selecting studies for systematic review: Inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders*, 33, 21–27. <https://pdfs.semanticscholar.org/a6b4/d6d01bd19a67e794db4b70207a45d47d82f3.pdf>
- Michie S, Thomas J, Johnston M, Mac Aonghusa P, Shawe-Taylor J, Kelly MP, ... Norris E (2017). The human behaviour-change project: Harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implementation Science*, 12(1), 121–133. doi: 10.1186/s13012-017-0641-5 [PubMed: 29047393]
- 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(6), 559–572. doi: 10.1080/13645579.2016.1252189
- Moher D, Liberati A, Tetzlaff J, Altman DG, & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*, 62(10), 1006–1012. doi:10.1016/j.jclinepi.2009.06.005 [PubMed: 19631508]
- Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, ... Stewart LA (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1. doi:10.1186/2046-4053-4-1 [PubMed: 25554246]
- Mortensen ML, Adam GP, Trikalinos TA, Kraska T, & Wallace BC (2017). An exploration of crowdsourcing citation screening for systematic reviews. *Research Synthesis Methods*, 8(3), 366–386. doi:10.1002/jrsm.1252 [PubMed: 28677322]
- Neimann Rasmussen L, & Montgomery P (2018). The prevalence of and factors associated with inclusion of non-English language studies in Campbell systematic reviews: A survey and meta-epidemiological study. *Systematic Reviews*, 7(1), 129. doi:10.1186/s13643-018-0786-6 [PubMed: 30139391]
- Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, & Worthy G (2008). The harvest plot: A method for synthesising evidence about the differential effects of interventions. *BMC Medical Research Methodology*, 8(1), 8. doi:10.1186/1471-2288-8-8 [PubMed: 18298827]
- Page MJ, McKenzie JE, & Forbes A (2013). Many scenarios exist for selective inclusion and reporting of results in randomized trials and systematic reviews. *Journal of Clinical Epidemiology*, 66(5), 524–537. doi:10.1016/j.jclinepi.2012.10.010 [doi] [PubMed: 23337785]
- Papageorgiou SP, & Biondi-Zoccai G (2016). Designing the review. In Biondi-Zoccai G (Ed.), *Umbrella reviews: Evidence synthesis with meta-reviews of reviews and meta-epidemiologic studies* (1st ed., pp. 57–80). Switzerland: Springer International Publishing.
- Pattanittum P, Laopaiboon M, Moher D, Lumbiganon P, & Ngamjarus C (2012). A comparison of statistical methods for identifying out-of-date systematic reviews. *PlosOne*, e48894. doi:10.1371/journal.pone.0048894
- Pieper D, Antoine S, Mathes T, Neugebauer EAM, & Eikermann M (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology*, 67(4), 368–375. doi:10.1016/j.jclinepi.2013.11.007 [PubMed: 24581293]
- Pieper D, Koensgen N, Breuing J, Ge L, & Wegewitz U (2018). How is AMSTAR applied by authors – a call for better reporting. *BMC Medical Research Methodology*, 18(1), 1–7. doi:10.1186/s12874-018-0520-z [PubMed: 29301497]
- Polanin JR, Hennessy EA, & Tsuji S (2019). Transparency and reproducibility of meta-analyses in psychology: A meta-review across 30 years of psychological bulletin. Submitted for Publication
- Polanin JR, Maynard BR, & Dell NA (2017). Overviews in education research: A systematic review and analysis. *Review of Educational Research*, 87(1), 172–203. doi:10.3102/0034654316631117
- Pollock A, Campbell P, Brunton G, Hunt H, & Estcourt L (2017). Selecting and implementing overview methods: Implications from five exemplar overviews. *Systematic Reviews*, 6(1), 145–158. doi:10.1186/s13643-017-0534-3 [PubMed: 28720141]
- Protogerou C, & Johnson B (2014). Factors underlying the success of behavioral HIV-prevention interventions for adolescents: A meta-review. *AIDS and Behavior*, 18(10), 1847–1863. doi: 10.1007/s10461-014-0807-y [PubMed: 24903669]
- Rosenthal R (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118(2), 183–192. doi:doi: 10.1037/0033-2909.118.2.183

- Sandieson RW, Kirkpatrick LC, Sandieson RM, & Zimmerman W (2010). Harnessing the power of education research databases with the pearl-harvesting methodological framework for information retrieval. *The Journal of Special Education*, 44(3), 161–175. doi:10.1177/0022466909349144
- Schmidt F, & Oh I (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes*, 121(2), 204–218. doi:10.1016/j.obhdp.2013.03.002
- Schünemann H, Bro ek J, Guyatt G, & Oxman A (Eds.). (2013). *GRADE handbook for grading quality of evidence and strength of recommendations* The GRADE Working Group Retrieved from <https://gdt.gradepro.org/app/handbook/handbook.html>
- Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, ... Kristjansson E (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358, j4008. doi: 10.1136/bmj.j4008 [PubMed: 28935701]
- Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, ... Bouter LM (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, 7, 10–17. doi:10.1186/1471-2288-7-10 [PubMed: 17302989]
- Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, & Moher D (2007). How quickly do systematic reviews go out of date? A survival analysis. *Annals of Internal Medicine*, 147(4), 224–233. doi:0000605-200708210-00179 [PubMed: 17638714]
- Tanner-Smith EE, & Tipton E (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, 5(1), 13–30. doi: 10.1002/jrsm.1091 [PubMed: 26054023]
- The Cochrane Collaboration. (2011). In Higgins JPT (Ed.), *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0 Retrieved from www.cochrane-handbook.org
- Thorne S, Jensen L, Kearney MH, Noblit G, & Sandelowski M (2004). Qualitative meta-synthesis: Reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, 14(10), 1342–1365. doi:10.1177/1049732304269888 [PubMed: 15538004]
- Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, ... Straus SE (2016). A scoping review on the conduct and reporting of scoping reviews. *BMC Medical Research Methodology*, 16, 15–016-0116–4. doi:10.1186/s12874-016-0116-4
- Valentine JC, & Cooper HM (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (study DIAD). *Psychological Methods*, 13(2), 130–149. doi: 10.1037/1082-989X.13.2.130 [PubMed: 18557682]
- Waddington H, White H, Snilstveit B, Hombrados JG, Vojtkova M, Davies P, ... Tugwell P (2012). How to do a good systematic review of effects in international development: A tool kit. *Journal of Development Effectiveness*, 4(3), 359–387. doi:10.1080/19439342.2012.711765
- Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, ... ROBIS group. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, 69, 225–234. doi:10.1016/j.jclinepi.2015.06.005 [PubMed: 26092286]

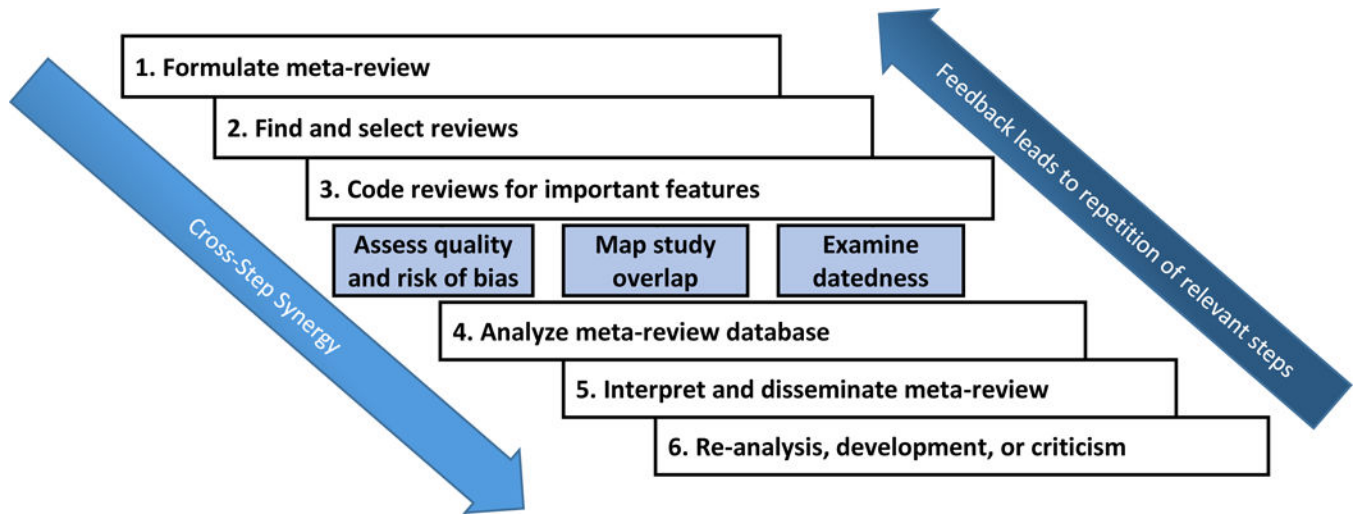


Figure 1. Flow diagram of best-practice methods in a meta-review from conceptualization to presentation of findings.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Meta-Review Strategies for Data Extraction from Primary Research Syntheses

Coding item	Description	Other notes
1. Was the review financially supported?	Three choices – “yes”, “no”, “not reported”; note whether a conflict exists.	Reporting requirements may vary by journal and discipline.
1a. If yes, by whom?	Could ask for narrative funding source or have categories for coders to select (e.g., government, university, non-profit, etc.).	
1b. Conflicts of interest	Typically, financial conflicts are reported, but other conflicts may be relevant (e.g., team members have published prior position papers on the effects of interest).	
2. Registered/published protocol	Best evidence that <i>a priori</i> planning took place.	If the protocol was registered and is available, it should be linked to the review for coding and assessment of bias purposes. The protocol should be compared to the published review for discrepancies between what was planned and conducted.
2a. If yes, note the location/hyperlink		
2b. If no, did authors mention creating one before beginning?	Provides some evidence of <i>a priori</i> planning.	
3. Review aims	Narrative description using authors' words	
4. Inclusion criteria of the review	Depending on scope of the meta-review, a series of questions broken into categories such as (1) population, (2) study design, (3) setting, (4) intervention type, (5) moderators, etc.	Useful for understanding population of studies in the overview and may be very helpful in grouping categories for synthesis
5. Number of eligible studies in review (overall), number of studies in specific analyses of interest and numbers of observations in each	Ultimately, observations of individuals are integrated in systematic reviews, meta-analyses, and meta-reviews.	Large sample sizes do not necessarily connote higher quality or more reliable findings.
6. Type of evidence synthesis	(1) Narrative systematic review; (2) Classical meta-analysis; (3) Individual-participant meta-analysis; (4) Network meta-analysis; (5) SEM meta-analysis; (6) RVE meta-analysis; (7) Multilevel meta-analysis; (8) Not reported; (9) None	
6a. Pooling method used	Given the diversity of approaches and term usages, review teams may wish to code this item descriptively and use the authors' terms and description of analysis approach.	
6b1. Qualitative synthesis approach	(1) Fixed effect; (2) Random effects; (3) Weighted-Least Squares; (4) Robust-Variance Estimation; (5) SEM; (6) Multilevel model	If there is non-independence between observations, then (4) - (6) are the most appropriate.
6b2. Quantitative modeling assumptions	Is it clear that the correct statistical information has been used to calculate effect sizes?	May only be applicable if including meta-analyses. Will need to plan how to handle missing or discrepant data.
7. Metric of the main effect(s) and other relevant effect size information, including complexities in ES calculation		
8. Heterogeneity statistics	Review authors should take heterogeneity into account in reaching their conclusions. Identify which heterogeneity statistics were reported by authors among the following: χ^2 and its associated <i>p</i> -value, τ^2 , I^2 . Note when reviews did not report these values or used them incorrectly in interpretations of trends.	Only applicable if reviews include meta-analyses. If moderator analyses were also conducted and of interest in the meta-review, additional heterogeneity statistics that should be collected: Adjusted R^2 , residual I^2 .
9. Was the review part of an update?	If it is an update, then collect all related reports for coding purposes.	Note any improvements of methods and determine whether the most recent version has used them even for studies summarized in earlier reports.

Coding item	Description	Other notes
10. Was publication bias /small study bias assessed?	Was heterogeneity taken into account in judging presence of bias?	Typically only applicable for meta-reviews of meta-analyses.
10a. If so, what type of assessment(s) were used?	This item could be an open text box or the typical tools used could be included, such as, funnel plots, Eggers or other regression-based tests, trim and fill, cumulative meta-analysis, fail-safe <i>N</i> , PET, PEESE, etc.	Although not every option included may be appropriated, it is still useful to capture what tool authors used. Note that some publication bias statistics offer advantages over others (e.g., (Becker, 2005)).
11. Was quality of included studies assessed in the meta-analysis?	Here, best to use answer options such as “yes” or “no” because authors should assume “no” if the author does not discuss study quality issues in the review.	If none judged quality, it is possible that substantive conclusions depend on studies with poor rigor.
11a. If so, what tools were used to assess study quality?	In addition to the standardized tools, may be useful to have options such as “author-created scale” or “single items.”	Even if meta-analyses use multiple-item inventories of quality, it is possible they have overlooked quality-relevant dimensions; studies satisfying all items may still have critical flaws.
11b. How did review authors examine the influence of study quality on outcomes?	In addition to using a study quality tool, reviews should examine whether key results depended on inclusion of studies with poor methods.	If meta-analyses assessed quality, did they use it to condition their findings?

Note. For some items, the coding sheet may simply need to reflect the uncertainty in coding that item by including coding options such as “clearly reported”, “not reported, but can be inferred”, and “not reported, cannot be inferred” or some such set-up. SEM=Structural equation model. RVE=Robust variance estimation.

Table 2.

Example content and steps for meta-reviews, synthesized (a) narratively and (b) quantitatively.

<i>a. Narrative Content Analysis for Meta-Reviews (generated from example provided by Protogerou & Johnson, 2014).</i>	
Review aim: Identify factors underlying success of behavioral HIV-prevention interventions for adolescents	
(1) Preparation	<ul style="list-style-type: none"> • Select appropriate/representative unit of analysis from the ‘universe’ • Data immersion
(2) Organization	<ul style="list-style-type: none"> • Open coding (iterative process) • Creating categories • Abstraction
(3) Reporting	<ul style="list-style-type: none"> • Report findings in a way that makes sense of the literature for the reader
b. Example “Semi-quantitative” Analysis for Meta-Reviews, Their Aims, and Analytic Steps.	
Citation(s)	Aim
(A) Chow & Ekholm (2018). (B) Polanin, Tanner-Smith, & Hennessy (2016)	Estimate extent of publication bias
(C) Ebrahim, Bance, Athale, Malachowski, & Ioannidis (2016)	Estimate bias due to industry involvement bias
(D) Johnson, Low, & MacDonald (2015)	Examine analytic practices of assessing and evaluating whether the quality of primary studies affects the results meta-analyses generate.
	<ol style="list-style-type: none"> 1. Extract data on funding sources and other key characteristics 2. Calculate proportions of reviews funded by different sources 3. Use exact tests to evaluate whether review conclusions differ by funding source 4. Report <i>ORs</i>, <i>95% CIs</i>, <i>p-values</i>
	<ol style="list-style-type: none"> 1. Extract data on primary study quality and other key characteristics 2. Calculate descriptive summaries 3. Calculate correlations