



Systematic and Nonsystematic Reviews: Choosing an Approach

8

David A. Cook

Overview

Systematic reviews and purposive (nonsystematic) reviews serve valuable and complementary roles in synthesizing the results of original research studies. Systematic reviews use rigorous methods of article selection and data extraction to shed focused, deep light on a relatively narrow body of research, yet of necessity may exclude potentially insightful works that fall outside the predefined scope. Purposive reviews offer flexibility to address more far-reaching questions and pursue novel insights, yet offer little assurance of a balanced perspective on the issue. This chapter reviews the strengths and weaknesses of each approach, and suggests specific questions to help researchers select among these approaches. Different approaches to quantitative and narrative research synthesis, including meta-analysis, are also described.

Practice Points

- Systematic and purposive (nonsystematic) reviews serve valuable and complementary roles.
- Thoughtful evidence synthesis is arguably the most important part of any review; both quantitative and narrative approaches are effective.
- In choosing a review approach, reviewers might ask: What is the purpose of the review? What is the current state of the literature? and, Which set of limitations matter more?

Health professions education research has shown tremendous growth in recent years, and with this comes an increased need for articles that synthesize the findings from individual

original research studies. Research syntheses (often called “review articles”) serve at least two distinct yet complementary purposes: they provide a succinct summary of what is known about a given topic, and they highlight gaps in our understanding that may warrant increased attention in future research.

Various labels are applied to reviews of different “types,” including systematic reviews, narrative reviews, critical reviews, scoping reviews, realist reviews, rapid reviews, and state-of-the-art reviews. Indeed, one group described 14 distinct review types [1]. However, I find such categories difficult to reliably discriminate in practice, not only because there are no universally-accepted definitions but because the boundaries overlap. For example, nine of these 14 review types previously mentioned were variations of a “systematic” review (e.g., qualitative systematic review, rapid review, and systematized review).

I prefer a simpler approach that classifies review articles as systematic or non-systematic (or, to use a less judgmental term, “purposive”). As will be elaborated later on, “systematic” reviews use a defined and reproducible approach in selecting articles and extracting data. Purposive reviews follow a more strategic and adaptable approach to selection and extraction. While some researchers disparage purposive (non-systematic) reviews, others criticize systematic reviews. Yet I believe that both systematic and purposive reviews have strengths and weaknesses that make them more or less appropriate depending on the researcher’s purpose or question. Another distinguishing feature of review types is whether they synthesize the original research findings using quantitative or qualitative methods; these distinctions are even more blurred than those for article selection and data extraction.

The goal of this chapter is to provide guidance to readers (i.e., would-be writers of reviews) on how to align their choice of review type and methods with their purpose. I will highlight a fundamental conceptual distinction between systematic and purposive reviews, present three questions to guide the selection among review types, describe approaches

D. A. Cook (✉)
Mayo Clinic Multidisciplinary Simulation Center, Office of Applied Scholarship and Education Science, and Division of General Internal Medicine, Mayo Clinic College of Medicine and Science, Rochester, MN, USA
e-mail: cook.david33@mayo.edu

to data synthesis, and conclude with a seven-step approach to planning a review that focuses on principles relevant to all review types. I will also touch briefly on three types of review that may contain elements of both systematic and purposive reviews, namely realist [2], scoping [3, 4], and state-of-the-art reviews.

Strengths and Limitations of Systematic and Purposive Reviews

Systematic reviews use predefined criteria for study inclusion and seek to extract the same information from each study, which usually includes a formal appraisal of study methodological quality. They often (but not always) use quantitative approaches to synthesis, which may include meta-analysis. The systematic approach identifies a comprehensive list of studies relevant to the research question, and distills presumably important information about each study. If done well, it defines the current state of research as regards chosen topic. It also helps to identify research gaps (e.g., populations or interventions notably absent among the studies found), characterize methodological strengths and deficiencies across studies, and avoid the bias that might arise from selecting only studies that support the author's preconceived position. The systematic approach would (in theory) allow another investigator to replicate the results and arrive at similar conclusions. However, reliance on a specific search strategy and distinct inclusion criteria prevents the systematic review from pursuing findings and ideas that are broadly relevant and potentially insightful but strictly fall outside the predefined scope. They are often perceived as narrow, sterile, and detached from the practical complexities of daily life. Systematic review are like lighthouses – they cast a powerful beam that illuminates the intended area of study, but leave the rest of the ocean in the dark. In addition, systematic reviews are not free of bias; every review involves countless decisions including those regarding the scope, search strategy, inclusion criteria, data selected for extraction, extraction process, and presentation of results.

Purposive reviews allow the researcher to reflect broadly upon a theme, drawing upon research, frameworks, and philosophy both within their field and from other fields (e.g., outside of health professions), and thereby yield insights that a systematic review could never achieve. Strategic selection of articles, unencumbered by the rules of the systematic review, further allows researchers to pursue ideas and findings that emerge unexpectedly during the process of the review, and to include a diverse spectrum of research methods. Discordant findings can be used to identify novel insights. Rather than comprehensively define the current state of evidence (what works), purposive reviews tend to address more far-reaching questions and generate novel

insights about why and how. However, there is no guarantee that the articles cited represent a balanced perspective on the issue; relevant work could have been inadvertently missed or even deliberately ignored. Purposive reviews act as a floodlight, illuminating a large area immediately near the source, but missing possibly important regions that lie farther away.

The strengths and limitations of systematic and purposive reviews parallel those of quantitative and qualitative research [5]. Both quantitative research and systematic reviews prefer large samples (of human participants or research studies) and emphasize systematic sampling. To minimize error, researchers seek that all subjects/studies be as similar as possible, and differences are viewed as error to be averaged out if possible. By contrast, both qualitative research and purposive reviews emphasize purposive, iterative sampling that shapes and is shaped by emerging insights. Rather than large samples, these approaches emphasize integrating information from multiple sources (triangulation). Differences between subjects/studies are viewed as opportunities to identify novel insights, often through extended data collection and new subjects/studies. Quantitative and qualitative research approaches are generally accepted as complementary, and the same should be true of systematic and purposive reviews.

Specific Review Subtypes

Realist Reviews and Scoping Reviews

Realist [2] and scoping [3, 4] reviews have received increased attention and use in recent years. Each employs a systematic albeit nonlinear approach to article selection and data synthesis, and can be considered a type of systematic review. However, they merit special attention because they have distinct purposes and defined methods, and some methods share features of purposive reviews.

The realist review was introduced as “a new method of systematic review designed for complex policy interventions” [2]. The realist approach is also well-suited for educational activities, which are typically complex. Realist reviews seek to elucidate the theoretical foundations of a given intervention or phenomenon, with particular emphasis on contextual influences (i.e., what works, for whom, in what context, and why). They are systematic in the sense that they use rigorous, transparent, and reproducible methods to search and synthesize the literature. However, a realist review explicitly involves a search for relevant theories and uses these theories to interpret the evidence found. Additionally, the search strategy and selection criteria typically evolve during the review and may include purposive elements [2].

Scoping reviews seek to provide a comprehensive snapshot of the literature in the field. They too are systematic in their use of rigorous, reproducible methods. However, in

contrast to the traditional systematic review in which the search strategy, inclusion criteria, data extraction items, and data analysis are largely pre-planned, each of these components typically evolves during the course of a scoping review. The authors identify up-front the scope of the review and define preliminary criteria for each component, but then add to and adjust each component as their understanding of the field grows. Scoping reviews may or may not report the actual results of any study, and sometimes include non-original-research literature such as editorials and other reviews. A well-done scoping review will identify and catalog the key terms, concepts, interventions, outcomes, and study designs extant in the field, thereby creating a map to guide future researchers and reviewers.

State-of-the-Art Reviews

State-of-the-art reviews provide an analysis of current work in the field, typically using a specific (recent) date as the criterion for inclusion (e.g., the last calendar year or past 5 years). They can otherwise adopt the methods of any other review type, with the corresponding strengths and weaknesses. The chief advantage is the emphasis on recent work, which is particularly important in a fast-moving field.

Options for Synthesizing the Evidence

All literature reviews extract evidence of some kind from the publications identified, such as numeric data, statistical test results, or themes. Synthesizing this evidence effectively is arguably the most important part of any review. I consider the synthesis approach separately from the review “type,” since both systematic and purposive reviews can appropriately use a broad spectrum of methods to synthesize and report their findings. Indeed, since text reporting and data visualization are inextricably linked with the synthesis process, there are an essentially infinite number of possible approaches to synthesis. Broadly speaking, however, synthesis approaches can be viewed as quantitative – presenting results as numbers; and qualitative – presenting results as a narrative (words). Whether quantitative or qualitative, data synthesis is an art that requires reviewers to put themselves in the shoes of the reader to anticipate and answer their questions, and provide relevant, succinct, and self-explanatory summaries and visualizations of supporting data.

Quantitative synthesis includes meta-analysis and a variety of other methods for reporting and integrating numeric data. Meta-analysis is simply a statistical technique that averages (“pools”) the results of several research studies, and estimates the magnitude of between-study differences (heterogeneity or inconsistency) that could signal important dif-

ferences in interventions, participants, settings, outcome measures, or study designs. Meta-analysis can also be used to examine, and hopefully explain, such inconsistencies. Although meta-analysis and systematic review are often colloquially viewed as interchangeable, in fact they are distinct. Many systematic reviews (likely a majority) use non-meta-analytic methods to synthesize results. Conversely, meta-analysis could, in principle, be applied to any review type; however, it is rarely employed in purposive reviews since most researchers would consider the results misleading in the absence of a systematic identification of studies (i.e., the pooled “best estimate of effect” would be inaccurate if any relevant studies were omitted). The question often arises: Is it appropriate to pool these results using meta-analysis? The answer always depends on the question asked; pooling across different populations (e.g., medical students and residents), interventions, outcomes, and study designs may or may not be appropriate depending on whether the resulting number makes sense and helps to answer the question. As I stated previously, “The most challenging aspect of conducting a meta-analysis ... is determining whether the original studies address a common question or framework. Analytic measures of inconsistency can help with this determination, but ultimately this is a conceptual – not a numeric – decision” [6]. Performing a meta-analysis does require skill with the statistical technique, but more important is to know what analyses are needed to support a meaningful and practical message.

Non-meta-analytic numeric synthesis can use a variety of tables, figures, and text to effectively report numeric data without pooling, for both systematic and purposive reviews. However, such reporting should emphasize the magnitude of effect (effect size) rather than the results of statistical tests. Effect sizes for different study designs include raw or standardized differences in scores, correlation or regression coefficients, and odds ratios. Reporting only the results of statistical tests (e.g., “Three studies found a statistically significant benefit.”) – so-called “vote counting” – is flawed for at least two reasons. First, vote-counting ignores the magnitude of effect: a large difference may be non-significant if the sample size is small, whereas even tiny differences will reach statistical significance with a large sample size. Second, it relies on a fixed notion of statistical significance (the P threshold of 0.05, while commonly used, is in fact arbitrary).

Most reviews – including systematic reviews and even meta-analyses – employ at least some features of qualitative (narrative) synthesis. Narrative synthesis is hard work! In addition to avoiding vote-counting, reviewers must not present a “litany of the literature” in which the results of each study are described in turn with only minimal integration. Rather, a good synthesis will first interpret and integrate the findings to reach a “bottom line” message that incorporates the strengths, weaknesses, inconsistencies, and gaps in the

evidence, as well as potential moderators such as populations, study designs, and contextual factors; and will then report this message together with a succinct summary of the evidence that supports the message. Narrative synthesis works with both numeric data and qualitative data.

Which to Use?

Deciding which type of review to employ depends on the answers to at least three questions.

First, and usually most important, *what is the purpose of the review?* Traditional systematic reviews address focused questions within a defined field. They seek to provide a comprehensive snapshot of current evidence within that field, including a bottom-line appraisal of “Does it work?” They typically identify areas in which evidence is lacking either from a paucity of studies, or from shortcomings in the available studies. Purposive reviews tend to address broader, far-reaching, and less defined questions. They seek to integrate findings across fields, often focusing on “Why or how it works?” in addition to the simpler “Does it work?” They likewise identify areas of needed research, but typically frame these as thematic deficiencies rather than limitations in the number or quality of studies. Some purposive reviews even redefine the question itself, refocusing or reframing our understanding of and research priorities for the field. Scoping reviews seek to present a snapshot of the published literature in a specified field. Realist reviews seek to understand the theoretical foundations for the selected intervention, with emphasis on contextual interactions (what works, for whom, in what context).

Second, *what is the current state of the literature?* Of course, answering this question is one of the reasons to do a review; but the researcher should have some sense of the answer. If there are lots of studies, and especially if the studies are of high quality and/or address very similar questions (e.g., the same type of intervention or the same population), then it might be reasonable to pursue a comprehensive listing and quantitative synthesis of these studies using a systematic review. Conversely, if there are few relevant studies, or the available studies reflect a variety of approaches, participants, interventions, or questions, then a purposive review might be more appropriate. In this case, the purposive review would allow the researchers to look beyond these few studies to identify work done in other fields, or work that addresses other questions and illuminates the topic even if not directly relevant. Of course, one could do a systematic review of very few studies, or a purposive review of a large body of evidence. A scoping review is helpful if the state of the field – the vocabulary, theories, interventions, outcomes, and overall volume of evidence – are truly unknown. A realist review requires a modest number of studies to explore the possible

contextual interactions, and compose meaningful evaluation of the underlying theories.

Third, *which set of limitations matter more?* Systematic reviews are limited by reviewers’ preconceived notions (biases) that may manifest in the inclusion criteria, the data selected for extraction, the processes of inclusion and extraction, the presentation of results, and the final conclusions. Adherence to the protocol prevents reviewers from pursuing interesting findings that fall outside the scope of the question and inclusion criteria. Most systematic reviews fail to accommodate the complexity of social interventions and interactions. Finally, the implicit trust that many readers naively render to systematic reviews could be viewed as a limitation. By contrast, purposive reviews suffer from clearly subjective inclusion criteria and data extraction, but at least they avoid any pretense of objectivity [7]. They also avoid the constraint of adhering to a protocol, and can more readily accommodate complexity. Scoping reviews are limited by incomplete appraisal of study quality, and by limited synthesis of evidence. Realist reviews are limited by the absence of quantitative synthesis and by the subjectivity encountered in identifying relevant theories and original research studies. Finally, both purposive and realist reviews require that the reviewers possess a fairly advanced understanding of the topic (i.e., to purposively identify relevant studies, theories, and conceptual frameworks), whereas in systematic and scoping reviews this understanding can develop over the course of the review.

Choosing the synthesis approach is dictated primarily by the needs of the emerging message. The message, and thus these needs, can be anticipated up front (e.g., if the purpose is to quantitatively summarize current evidence, then a meta-analysis may be required). However, in many cases the ideal synthesis approach evolves as reviewers examine the accumulating data and contemplate how best to share the insights they are discovering. For example, the data might simply not support a planned meta-analysis, or a graphical representation of numeric data may be added to complement a planned narrative synthesis. As I conceive the tentative synthesis approach during the planning stage, I typically write out a rough draft of the Results section, including sketching key tables and figures.

Two often-cited considerations should *not* be part of these decisions: time and team size. All of these review types require a substantial investment of time to do well; none of them should be viewed as a “fast track to a publication.” Time will be determined more by the volume of literature reviewed than the review type per se. All of these review types also require a team approach; a minimum of two reviewers, and often substantially more, is required in conducting a high-quality review to avoid systematic bias, minimize random error, deepen insights, enhance interpretation, and distribute the workload.

A Seven-Step Approach to Planning a Review

In closing I will share seven tips for planning a review of any type, based on points I outlined previously [8].

Clarify the Question

All research projects begin with a clear question or purpose, and literature reviews are no exception. As acknowledged above, questions can differ widely, variously focusing on identifying problems, clarifying theory, testing theory, quantifying impact, or mapping the current state of the field. Borrowing from a framework first proposed for original research studies [9], a review's purpose "might be classified as description (historical or descriptive overview), justification (synthesis of evidence to identify the current state with weak reference to a conceptual framework), or clarification (synthesis of evidence to understand mechanisms, identify gaps, and build a conceptual framework)" [8]. Although all these purposes have merit, clarification studies tend to advance our understanding more than descriptions or justifications [9].

Pick an Approach That Matches the Question

Once the question has been identified, reviewers must select the review type most appropriate to answer that question. As outlined above, these decisions revolve around systematic vs. purposive approaches to study identification and data extraction, and quantitative vs. qualitative approaches to data synthesis.

Plan Defensible Methods

Ideally, reviewers will develop and follow a written plan for conducting the review. The planned methods will depend upon the question and the selected review type. Methods for systematic reviews have been described in books [10, 11], journals [12, 13], and online resources [14, 15], and *reporting* guidelines like the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [16] (PRISMA) highlight important methodological considerations. Guidelines for scoping [3, 4] and realist [2, 17] reviews have also been outlined. By contrast, purposive reviews are much more flexible, and as such do not have universal standards. However, principles of high-quality *original* qualitative research can provide guidance in conducting a high-quality qualitative *literature synthesis* as well; these principles include clarification of purpose, recognition of researcher assumptions and perspectives (reflexivity), working in research teams, purposeful sampling, thoughtful analysis that makes a conscious effort to consider alternate perspectives, and detailed presen-

tation ("rich description") of evidence that both supports and counters the bottom line message.

Set the Stage

Just as with original research, the Introduction should set the stage for the review by summarizing relevant literature to justify the need for a review on this topic, and to clarify relevant theories and frameworks. In justifying the need, reviewers should highlight the strengths and shortcomings of relevant previous reviews rather than citing original research, since relevant original research studies will typically be identified during the review and then cited in the results. Shortcomings in previous reviews do not necessarily arise from methodological weaknesses; they can also arise from differences in their scope, age, type, and synthesis approach. The Introduction should clarify how these shortcomings leave an important gap in our understanding, and how the proposed review will fill this gap.

Organize and Interpret to Share a Clear Message

Reviewers often focus their efforts on identifying and selecting studies and extracting information from them. However, it is equally important – and often more challenging – to effectively synthesize the results of these studies into a meaningful and well-supported message. A review is only as good as its bottom line message; the method of synthesis is a vitally important means to that end.

Appraise Study Quality and Explore the Impact on Review Conclusions

Depending on the review question and scope, the evidence collected and reviewed might take many forms – randomized trials, non-randomized experiments, correlational studies, surveys, assessment validation studies, and various forms of qualitative research. Each of these study designs has "best practice" features that, if followed, strengthen our confidence in the results. These features should be taken in account when drawing conclusions from the data synthesis. The appraisal of study quality is a formal part of most traditional systematic reviews; in all other review types, the quality appraisal should be equally thorough albeit perhaps less formal. Importantly, it is not enough to just describe or enumerate the various design features (as I have often seen done). Rather, the strengths and weaknesses of a given study should determine the degree to which those results influence the bottom-line conclusions of the review. Finally, although method quality checklists have been developed for many study designs, it is the specific

design feature (e.g., randomization, loss to follow-up, blinded assessment) not a total quality score that should be emphasized in conducting this integration. The relative importance of one design feature over another may vary for different reviews; a rote one-size-fits-all approach is discouraged [18].

Report Completely

Reviewers must present a complete and transparent report of what they did and what they found. In reporting their methods, reviewers should describe in detail what they did and the key decisions they faced. There are no “standard procedures” for any review type, and nothing can or should be taken for granted. In reporting their findings, they should describe in detail both the processes of the review (such as inter-rater agreement on inclusion or extraction, the source and number of studies considered and included, or the conceptual frameworks considered while interpreting results) and the methods and results of the included studies. Reporting guidelines (not to be confused with method quality appraisal tools) such as those for systematic [16], realist [17], and scoping [3, 4] reviews can, when available, remind reviewers what information to report. Limitations both in the review methods and in the number, quality, and relevance of the original research studies will influence the findings of the review. These limitations should be acknowledged and, as noted above, accounted for in formulating the synthesis.

References

- Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf Libr J*. 2009;26:91–108.
- Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review – a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy*. 2005;10(Suppl 1):21–34.
- Arksey H, O’Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8:19–32.
- Levac D, Colquhoun H, O’Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010;5:69.
- Cook DA. Narrowing the focus and broadening horizons: complementary roles for nonsystematic and systematic reviews. *Adv Health Sci Educ Theory Pract*. 2008;13:391–5.
- Cook DA. Randomized controlled trials and meta-analysis in medical education: what role do they play? *Med Teach*. 2012;34:468–73.
- Eva KW. On the limits of systematicity. *Med Educ*. 2008;42:852–3.
- Cook DA. Tips for a great review article: crossing methodological boundaries. *Med Educ*. 2016;50:384–7.
- Cook DA, Bordage G, Schmidt HG. Description, justification, and clarification: a framework for classifying the purposes of research in medical education. *Med Educ*. 2008;42:128–33.
- Cooper H, Hedges LV, Valentine JC. *The handbook of research synthesis*. 2nd ed. New York: Russell Sage Foundation; 2009.
- Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley-Blackwell; 2008.
- Cook DA, West CP. Conducting systematic reviews in medical education: a stepwise approach. *Med Educ*. 2012;46:943–52.
- Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med*. 1997;126:376–80.
- Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions* 5.1.0 [updated March 2011]. Available at: <http://handbook-5-1.cochrane.org/>. Accessed 31 May 2018.
- Campbell Collaboration. Campbell Collaboration Resource Center. Available at: <https://campbellcollaboration.org/research-resources/research-for-resources.html>. Accessed 31 May 2018.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann Intern Med*. 2009;151:264–9.
- Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. *BMC Med*. 2013;11:21.
- Montori VM, Swiontkowski MF, Cook DJ. Methodologic issues in systematic reviews and meta-analyses. *Clin Orthop Relat Res*. 2003;413:43–54.