

OCTOBER 2012

Supporting Policy and Program Decisions: Recommendations for Conducting High Quality Systematic Evidence Reviews

Susan Zief and Roberto Agodini

Policy and program decisions typically involve selecting one choice from among a set of options, and research about the effect of those options can help inform the decision process. However, for the research to be useful, decision makers need a way of drawing accurate lessons from what often can be a large assortment of relevant studies. Systematic reviews can be particularly useful in this process because they identify, assess, and synthesize key pieces of evidence on policy or program effectiveness. This brief provides recommendations for conducting high quality systematic reviews. We hope that the recommendations will increase the number of such reviews, to provide decision makers with a greater number of useful evidence summaries that can inform decision making.

Systematic reviews are a useful tool for decision makers because they identify relevant studies about a policy or program of interest, and summarize the findings across the various studies. A review is “systematic” when it follows predefined, transparent processes and standards that allow readers to understand the basis of the summary findings. Following such processes and standards also makes it easy to replicate or supplement the review at a later date. High quality systematic reviews define their processes and standards in such a way that the summarized findings are accurate, meaning that they are free from bias that could be introduced through the processes that guided the review effort or the individual study designs.

Since the 1993 founding of the Cochrane Collaboration, medical

practitioners and health policymakers have had an objective and trusted resource for systematic reviews of medical and health policy research. Social policy researchers have also begun similar efforts. The international Campbell Collaboration, the U.S. Department of Education, the U.S. Department of Health and Human Services, the U.S.

Department of Justice, and the Department for International Development in the United Kingdom are all supporting objective efforts to identify, assess, and synthesize effectiveness evidence. In 2011, the Institute of Medicine released a comprehensive set of guidelines for conducting systematic reviews, intended to standardize and

Recommendations for Conducting High Quality Systematic Reviews

1. Fully understand the goals and options to be covered by a systematic review.
2. Use best-practice literature search methods and clearly describe the approach up front.
3. Modify the approach only if it would increase the usefulness of the summary findings.
4. Follow established scientific standards to assess the quality of the studies.
5. Synthesize the findings in a way that is accessible to the intended audience.

improve the quality of review methods and procedures.

Our recommendations for conducting systematic reviews are based on a simple concept: the usefulness of a review is only as good as the relevance and quality of the evidence included and the approach used to synthesize it. The recommendations are intended to be an introduction to the elements of high quality systematic reviews, rather than a detailed description of how to conduct them.

Recommendation 1: Fully understand the goals and options to be covered by a systematic review

Systematic reviews are often framed by policy and program questions. To answer those questions, choices need to be made. The purpose of a systematic review is to support the decision making process by synthesizing research about the effect of the policies or programs of interest. A plan for conducting a systematic review should identify the relevant choice set for the policy and program question: What outcome do policymakers want to affect, for whom, and with what types of interventions?

The research questions that initiate systematic reviews may originally be phrased broadly. Do after-school programs improve student outcomes? Are teen pregnancy prevention programs effective? Do charter schools make a difference? The first stage of a review involves identifying the interventions that are relevant to the question. For example, a review focused on after-school programs would begin by defining such programs—do they include anything available for students during the after-school hours, or only more specific models, such as tutoring, mentoring, or academically focused programs? Will a teen pregnancy prevention review consider everything that might possibly have an impact on

preventing teenage pregnancies (including broader youth development strategies), or focus on programs that teach abstinence or other approaches for delaying intercourse and protecting against pregnancies? Will a charter school review focus on a broad strategy, such as the introduction of charter schools, or on a particular charter management organization?

Another important consideration is the target population and context the review will encompass. For example, some reviews may be limited to interventions studied with individuals from particular socioeconomic backgrounds. Others may be limited to interventions studied in a particular geographic context, such as urban areas.

The systematic review should consider and define the outcomes of interest to decision makers. For example, to answer the question “Does educational technology improve student outcomes?” it is critical to identify the outcomes that are relevant to the question. Is the goal to understand how such technology affects academic achievement (scores on math or reading assessments), computer skills, and/or behavioral indicators (school attendance)?

These decisions about the criteria used to identify studies that are relevant for the review—criteria about the type of intervention, target population, context, and outcomes examined—are important because they define the scope of the review, which affects the time and resources needed to conduct the review and the conclusions that can be drawn. A narrow scope tends to limit the amount of relevant research for a review, reducing the effort needed to assess and synthesize it, but also may limit the conclusions that can be drawn in undesirable ways. In contrast, a broadly cast net with few specifications tends to capture evaluations of diverse program models, populations,

contexts, and outcomes, which could make it difficult to draw conclusions from the review.

To conduct a high quality systematic review, there must be sufficient resources available to support the intended scope. If that scope cannot be supported with the available resources, the review can be conducted but the summary findings should clearly identify any options that were not included in the review. Identifying the options that were not included also indicates how the review could be updated if the need arises and resources are available in the future.

Recommendation 2: Use best-practice literature search methods and clearly describe the approach up front

For a review to be of the most use, all research evidence—both published and unpublished—should be included and a clear protocol to guide the search must be established. Published research, which is easily found and downloaded using well-structured keyword searches in reference databases, is known to be biased toward positive findings. Identifying and acquiring copies of unpublished, or “gray,” research—such as dissertations or papers available only on program or research organizations’ websites—requires more resources. Reviewers need to make these efforts, including establishing a clear protocol for the search, for the review to be fully comprehensive. If resources are limited, the search strategy can be designed to capture the more readily available research, and the review can then be viewed as a solid foundation to be updated as resources allow. Any limitations of the review findings that could be attributed to the process of selecting studies for review (such as only including published journal articles) should be stated clearly.

Designing an efficient search strategy that can be executed with available resources is a potential challenge. It is a good idea to work with a reference librarian to identify key terms for the database search. Manually searching journals, which is extraordinarily labor intensive, often yields useful information only for months not already captured in a database search. General internet searches tend to yield little relative to the effort required to track down all the hits. Instead, it is advisable to identify organizations that produce relevant research and conduct targeted searches of their websites. It also is important to clearly describe the strategy before the search begins. Defining the strategy will create a road map for those conducting the search, which will help ensure the review is objective and that the summary findings are valid and useful.

**Recommendation 3:
Modify the approach only if it would increase the usefulness of the summary findings**

The search process may unearth a study that provides support for a different way of meeting the policy goal than the options considered when initially establishing the scope for the review. Alternatively, someone outside the review team may tell the team about a relevant study that was not identified through the search process.

If these situations arise, the reviewers and decision makers should ask: Would modifying the relevancy criteria or search process increase the usefulness of the findings and maintain the objectivity of the review process? Because such a modification would involve systematically expanding the approach, it also is important to ask: Are sufficient resources available to support the modification?

If the answer to both questions is “yes,” it would be useful to modify the approach because it would enhance the usefulness of the review. However, if the answer to either question is “no,” it is best to adhere to the initial plan to (1) refrain from making any modifications that might threaten the objectivity of the review (and the validity and usefulness of the findings) and/or (2) keep the cost of the review within the available resources.

**Recommendation 4:
Follow established scientific standards to assess the quality of the studies**

If the objective of a review is to understand the impact of a program or policy, it is important that the designs of the included studies can support claims of program effectiveness. Experimental (or random assignment) studies provide the most credible evidence of impacts. Some systematic reviews also consider evidence from quasi-experimental studies. For either type of design, seemingly small decisions made about the study design, its implementation, and the analyses can threaten a study’s integrity and the extent to which the findings are believable. For example, an experimental study’s ability to support causal statements is compromised if a significant fraction of its participants exit the study, particularly if the attrition rate differs across the treatment and control groups.

Studies should be included in systematic reviews only after the scientific quality of the evidence is assessed against rigorous standards, and it is determined that the studies can support claims of effectiveness. Fortunately, standards to assess a study’s design, implementation, and analytic decisions already exist, so there is no need to develop them

from scratch. For example, the education research community has helped create scientific standards for assessing evaluation studies, such as the What Works Clearinghouse evidence standards. A similar set of standards guides the U.S. Department of Health and Human Services’ Pregnancy Prevention Research Evidence review.

Even when standards to assess study design, implementation, and analyses are carefully defined, decisions about study quality are often complex and require the judgment of trained and experienced reviewers. The key elements of a strategy to ensure consistent application of rigorous standards include trained reviewers, the use of multiple reviewers, and an effective quality assurance process. Ideally, the following procedures should be put in place:

- Training and testing reviewers on how to apply the standards
- Reviewing each study by more than one trained reviewer
- Implementing a system of checks and balances including quality assurance by a trained senior researcher, to ensure that the reviewers do not make similar errors in applying the standards.

The review findings should be carefully documented and made publically available to ensure the transparency of the review process.

Incomplete reporting by study authors also can make it difficult to draw a clear conclusion about the quality of some studies. Ideally, reviewers can contact the authors to clear up any ambiguities or resolve questions raised by incomplete reporting, but that is not always possible. In these cases, we recommend a conservative approach that resists the temptation to make assumptions or give the study the benefit of the doubt.

Recommendation 5: Synthesize the findings in a way that is accessible to the intended audience

The audience for systematic reviews often includes practitioners and policymakers who need to make decisions based on sound evidence of what works. As such, the summary findings must simplify complexities with respect to varying interventions, outcomes, populations, contexts, and study designs. However, this user-friendly approach should be careful not to mask variation in the findings that is important for policymakers and practitioners to understand.

Summary findings should be clear about what is being studied and for whom, as well as the degree of confidence in the findings. Consider showing the findings by policy option or program model, as well as by outcome and quality of the evidence. Translate technical estimates of impacts into terms that are accessible and understandable for the audience, such as a metric that shows the degree of improvement relative to the outcome that would have occurred in the absence of the intervention. Assess the breadth and depth of the evidence so that decision makers can better understand the degree to which the findings may be representative of broader contexts and populations.

Basis for the Recommendations in this Issue Brief

The recommendations presented in this brief are based on extensive experience conducting systematic review efforts over the past 12 years. In 2000, Mathematica researchers conducted reviews of effectiveness evidence on after-school programs for the Campbell Collaboration and produced technical methods papers that influenced standards for assessing study quality. Our work on the What Works Clearinghouse of the U.S. Department of Education over the past decade contributed to the creation of evidence standards used to judge the quality of effectiveness evaluations in education and has resulted in the production of many systematic reviews across a broad range of topical areas in education. Most recently, our portfolio has expanded to include contracts from the U.S. Department of Health and Human Services to review evidence on early childhood home visitation, teen pregnancy prevention, and responsible fatherhood and family strengthening models. In addition, we have reviewed the effectiveness of interventions to improve outcomes for populations with barriers to employment for the nonprofit firm REDF.

Final Thoughts

It is possible that little or no research will be found that is relevant and of high enough quality to include in a review. In these situations, the existing research does not provide an accurate answer to the policy or program question. Stating this clearly is preferable to changing course or accepting lower quality evidence. This review outcome directs attention to gaps in the research base, which might encourage funders to support evaluation studies that address questions of interest to decision makers.

At the end of the review process, it is natural for reviewers to want to be able to draw definitive conclusions about effectiveness. But a systematic review cannot absolutely affirm what is and is not effective—rather, it can only convey what the available evaluation studies show. The result of a review is not to tell decision makers what to do, but rather to state what the field knows based on the existing evidence.

For more information, contact CIRE@mathematica-mpr.com.