

ASPE RESEARCH BRIEF

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION
OFFICE OF HUMAN SERVICES POLICY - U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

USING SYSTEMATIC REVIEWS TO INFORM POLICY INITIATIVES

Lessons from the HHS Teen Pregnancy Prevention Evidence Review

In fall 2009, the U.S. Department of Health and Human Services (HHS) launched a systematic review of the research literature on programs to prevent teen pregnancy, sexually transmitted infections (STIs), and associated sexual risk behaviors. Findings have been used to inform two new federal policy initiatives aimed at supporting evidence-based approaches to teen pregnancy prevention. This research brief highlights six key lessons from the review, intended to help the growing number of other organizations and federal agencies considering similar policy-driven reviews.

Recent years have seen growing interest in the use of systematic reviews to inform new policy initiatives, especially at the federal level. In fall 2009, HHS launched a systematic review to identify evidence-based home visiting programs for support under the Maternal, Infant, and Early Childhood Home Visiting Program of the Patient Protection and Affordable Care Act (Avellar and Paulsell 2011). Around the same time, HHS also launched a systematic review of programs to reduce teen pregnancy, STIs, and associated sexual risk behaviors, to help guide two new federal funding initiatives: the Office of Adolescent Health Teen Pregnancy Prevention (TPP) program and the state Personal Responsibility Education Program (PREP).

Such reviews offer many potential benefits: grounding new policy initiatives in scientific research evidence, directing scarce public resources to programs with the strongest available evidence of effectiveness, and stimulating future research and program development by creating incentives for effective programs and rigorous research.

But they also bring new challenges. Conducting a high quality systematic review can require significant time and

ABOUT THIS RESEARCH BRIEF

This ASPE Research Brief on the use of systematic evidence reviews in policy initiatives was written by Brian Goesling of Mathematica Policy Research. Since 2009, Mathematica and its partner, Child Trends, have conducted the Teen Pregnancy Prevention Evidence Review under contract with HHS. The review aims to identify, assess, and rate the rigor of program impact studies on teen pregnancy and STI prevention programs.

Office of the Assistant Secretary for Planning and Evaluation

Office of Human Services Policy

US Department of Health and Human Services

Washington, DC 20201



resources, and the ultimate value of the review findings for policy depends on the quality of the supporting research evidence. Clear policy direction results only if the review uncovers a strong and consistent body of evidence.

Lesson 1: Don't Reinvent the Wheel

Many resources are now available for planning a new systematic review. For example, the Institute of Medicine (IOM) recently released a set of comprehensive guidelines for systematic reviews, covering everything from literature search methods to analysis and reporting (IOM 2011). The IOM guidelines were not available when this review of teen pregnancy prevention programs began in 2009. However, the review team achieved similar results by borrowing from the standards and procedures used by existing systematic reviews and evidence assessment groups, such as the [Cochrane Collaboration](#), [Campbell Collaboration](#), [Blueprints for Violence Prevention](#), the [Centers for Disease Control and Prevention HIV/AIDS Prevention Research Synthesis](#), the [National Registry of Evidence-Based Programs and Practices](#), and the U.S. Department of Education's What Works Clearinghouse ([WWC](#)).

Drawing on such existing resources is the easiest and most efficient way to start a new review, especially for projects on a short time line or with limited resources. It also helps to establish the credibility and rigor of the review by grounding it in accepted standards and procedures. By drawing on existing resources, researchers also help promote greater consistency across systematic reviews, which should ultimately make review findings easier for the public to interpret.

These recommendations do not imply that one size fits all. Although the standards and procedures developed for this review of teen pregnancy prevention programs drew on those used by existing reviews and evidence assessment groups, they still had to be tailored to the specific goals of the project, the project's time line and resources, and the unique characteristics of the teen pregnancy prevention literature. The review team could not rely solely on existing resources to determine which outcome measures to examine, the range of program models to consider, or which journals or databases to include in the literature search. No two reviews are exactly alike, so some level of customization will always be needed.

ABOUT THE HHS TEEN PREGNANCY PREVENTION EVIDENCE REVIEW

Conducted for HHS by Mathematica Policy Research and its partner, Child Trends, the review aims to identify, assess, and rate the rigor of program impact studies on teen pregnancy and STI prevention programs. The review is an examination of research evidence, not program content or approach. The review assesses the quality and rigor of each included study, without respect to program content or delivery method. An initial review of the evidence was conducted from September 2009 through April 2010 and released in conjunction with the Office of Adolescent Health TPP program grant announcements. The findings were also highlighted in the 2010 state PREP grant announcement. The review findings were updated in April 2012, with plans for periodic updates in the future.

Lesson 2: Take Care in Defining the Scope of the Review

A main challenge encountered in planning the review involved a seemingly basic task: defining the range of program models to study. The review initially aimed to cover the full universe of teen pregnancy prevention programs. Defining the universe of programs was difficult, however, because the research literature provides no clear guidance on where to draw the line. The most common teen pregnancy prevention programs feature classroom-based curricula delivered to middle or high school-age students. But studies suggest that a broad range of other types of program models, from early childhood education to broader youth development and service learning programs, can also shape teen pregnancy and associated risk behaviors.

Tying the scope of the review to an external benchmark, such as the particular policy issue or decision the review aims to inform, can be a more effective approach. In the case of teen pregnancy, this meant aligning the scope of the review with the range of program models eligible for funding under the new federal TPP program. The literature search was kept as broad and expansive as possible. But then when deciding which particular studies to include in the review, the focus was limited to only those programs falling within the scope of the TPP funding announcement, which included curriculum-based programs and youth development approaches. This strategy both avoided the need to define the full universe of program models and helped link the review to the motivating policy initiative.

Other reviews have addressed this challenge by limiting their scope to a defined list of program models. For example, both the WWC and Home Visiting Evidence of Effectiveness reviews begin by prioritizing a specific list of program models to include in the review. The review team then focuses on identifying and assessing studies relating only to the selected program models. As long as the programs are selected in advance and without respect to the study findings, this approach presents little risk of bias and avoids the challenge of drawing a line around the full universe of program models.

DEFINING FEATURES OF SYSTEMATIC REVIEWS

Systematic reviews differ from more informal literature reviews or evidence summaries in two key ways: rigor and transparency.

Rigor. Systematic reviews follow a defined set of rules for identifying and assessing individual studies and synthesizing a body of evidence. The expert opinions or judgments of the review team may factor into the initial definition of these rules. Once established, however, the rules—not the opinions or judgments of the review team—determine the outcomes of the review.

Transparency. Users of a systematic review should have the ability to replicate the review findings on the basis of the standards and procedures described in the review protocol. No review is perfect, and critics may question the choice of certain standards and procedures over others. But as long as the review protocol states these decisions clearly, the integrity of the review remains intact.

Lesson 3: Engage Study Authors and Outside Experts

As a part of this review of teen pregnancy prevention programs, the review team has made several efforts to involve study authors and other outside experts in different stages of the review: distributing a public call for studies; giving professional conference presentations; hosting a public webinar; organizing a one-day panel meeting with experts in research methodology and systematic reviews; and corresponding with study authors to request clarifying details or additional information about studies under review.

These efforts to engage study authors and outside experts have demanded significant time and resources. The call for studies has required developing a system for distributing the call and receiving electronic submissions, monitoring and processing submissions, and responding to questions from study authors. Each time a new call has been released, about six weeks have been provided for responses, lengthening the project schedule. To receive outside comment on the review, the review team has had to identify or create venues to present the review findings and devote staff time to preparing and giving public presentations. Corresponding with study authors has been especially time- and resource-intensive, as each contact must be individualized to the particular author and study. Corresponding with study authors has also lengthened the review schedule, as it is unrealistic and burdensome to demand immediate responses.

Despite these challenges, the process of seeking expert guidance and contacting study authors has been worth the effort. The call for studies has helped to identify new or unpublished research—important for avoiding the risk of reporting bias that results from focusing only on published journal articles (IOM 2011). Public presentations and expert consultations have provided useful feedback on the review and helped increase the transparency of the review process. Corresponding with study authors has proven particularly important, as published research articles often lack the level of detail needed for systematic reviews. Even when the requested information is unavailable or an author is unresponsive, the process of asking for additional information and detail helps make the review as thorough as possible.

Lesson 4: Report on More than Just Positive Findings

A main goal of this review is to identify programs with evidence of effectiveness in reducing teen pregnancy, STIs, and sexual risk behaviors. The review findings thus highlight a list of programs meeting the review criteria for study quality and evidence of effectiveness. For the programs on this list, the review findings highlight the specific outcomes the program has been shown to affect: teen pregnancy, STIs, or associated sexual risk behaviors. But the review also reports null findings—that is, outcomes tested but found *not* to show evidence of favorable program effects. In addition, the review reports findings for programs that did *not* make the evidence-based list, including the program name, study citation, and a brief explanation of why the study fell short. For the few program models that have been evaluated more than once (see box on next page), the review reports on all prior studies of the program, not only those reporting positive effects.

This type of full reporting is an expected standard of systematic reviews—both to increase the transparency of the review process and to ensure an accurate interpretation of the review findings (IOM 2011). Full reporting also makes the review findings useful for different audiences.

Whereas some users are interested only in the programs that meet the review criteria, others want to know why a particular program model or study did not meet the criteria. Full reporting also helps to uncover gaps in the literature and directions for future research. For example, about half of the studies in this review of teen pregnancy prevention programs failed to meet the review criteria for study quality, identifying a general need for improved research quality beyond the specific findings reported for any one particular program model.

To help document these findings, the review team assessed each study in two steps. First, the team examined each study for methodological quality and rigor, focusing especially on the risk of bias in the study's impact estimates. For studies passing this quality bar, the team then made a second assessment, examining the impact findings for evidence of favorable effects. This two-step process enables users of the review to better understand why certain programs fell short of the review criteria. Some programs were excluded because they failed to meet the review criteria for study quality, whereas others met the quality standards but did not show evidence of favorable program effects.

Lesson 5: Expect that Some Review Findings Might Not Be Implementation-Ready

In using systematic reviews to inform new policy initiatives, program offices and sponsors should prepare for the possibility that not all programs, policies, or practices identified as effective by the review will be ready for replication on a wider scale.

Systematic reviews are good for assessing the quality and strength of a body of research evidence, but they generally do not account for the content of a particular policy, program, or practice, or how well or quickly it can be implemented. For this review of teen pregnancy prevention programs, some program models that met the review criteria had most or all of the necessary training and materials available for immediate implementation. Others, however, were still under development or had no formal materials available. In other cases, program materials were available but outdated—either factually or in the cultural references used in the materials.

There are two potential ways to address this challenge. One is to provide time for assessing program content and implementation readiness after release of the initial review findings. This assessment might draw in part from information collected during the review, but it also likely requires additional contact with program developers to collect more detailed information on

SUMMARIZING FINDINGS ACROSS STUDIES

Most teen pregnancy prevention programs have been evaluated only once, so the review has so far avoided the challenge of having to combine or summarize findings across multiple studies of a single program or intervention. Other reviews have addressed this challenge by conducting a formal meta-analysis—that is, statistically averaging program impacts across multiple studies. Another approach is to simply count the number of studies showing positive, null, or negative effects—the vote counting method of summarizing review findings. Review authors also face the challenge of determining when two studies are similar enough to combine or when they must be kept separate. Currently, there are no universal standards for addressing these challenges, so review authors must address them separately for each new review.

implementation experience and available training and materials. As an alternative approach, the initial review could be limited to only those programs, policies, or practices deemed ready for implementation. This second approach would require additional work during the initial planning and screening stages of the review, but it would save time later when the review findings are released.

Lesson 6: Use the Review Findings to Encourage Improved Research Quality and Reporting

One useful byproduct of this review has been a comprehensive and objective assessment of the relative strengths and weaknesses of the teen pregnancy prevention literature. On the plus side, the review findings show that the evidence base on such programs has grown substantially in recent years, and that researchers have had notable success in implementing randomized controlled trials with a range of different program models and in diverse settings. But the review findings have also uncovered areas needing improvement. Many teen pregnancy prevention studies provide incomplete information on study design and execution, effect sizes are often missing for key outcomes, more than half the studies reviewed did not meet the review standards for methodological quality, and few program models have been subject to replication studies.

Such findings can be used to encourage improved methodological quality and reporting standards in future research. At the simplest level, disseminating the review findings in public reports, journal articles, and conference presentations can make researchers more aware of current gaps in the field and shape the direction of future research. Stronger incentives can be achieved by adding new standards or requirements to future updates of the review. For example, to address the problem of incomplete reporting or missing effect sizes, future review updates could specify a minimum set of reporting requirements to qualify for review. To maximize these incentives and keep the review process fair and transparent, any changes to the review standards should be announced before they are implemented and should be clearly documented in the review protocol.

Summary and Conclusions

Lessons from the HHS review of teen pregnancy prevention programs suggest that systematic reviews can play an important role in shaping the direction of new programs and policy initiatives. Several keys to success are grounding the review in existing guidelines and standards, defining the scope of the review as clearly as possible, engaging study authors and other outside experts in key stages of the review process, and reporting the full range of review findings, not just evidence of positive effects. Review sponsors should prepare for the possibility that not all program models or practices meeting the review criteria will be ready for immediate implementation. They should also seek to use the review findings to encourage improved research quality and reporting standards. All systematic reviews will encounter unforeseen challenges, but they offer the best available method for synthesizing a body of evidence and grounding new policy initiatives in scientific research.

References

Avellar, Sarah and Diane Paulsell. "Lessons Learned from the Home Visiting Evidence of Effectiveness Review." Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2011.

IOM (Institute of Medicine). "Finding What Works in Health Care: Standards for Systematic Reviews." Washington, DC: The National Academies Press, 2011.

**DEPARTMENT OF HEALTH
& HUMAN SERVICES**

Office of the Secretary
Washington, DC

OFFICIAL BUSINESS
Penalty for Private Use \$300

