



November 2023

Understanding the Teen Pregnancy Prevention Evidence Review

In this brief, we provide an overview of the U.S. Department of Health and Human Services (HHS) Teen Pregnancy Prevention (TPP) Evidence Review, an ongoing systematic review of the teen pregnancy prevention literature designed to identify programs and program components with evidence of effectiveness in reducing teen pregnancy, sexually transmitted infections (STIs), and associated sexual risk behaviors. This brief is intended for researchers planning or implementing an evaluation of a TPP program or program component, to provide information about the review process and requirements.

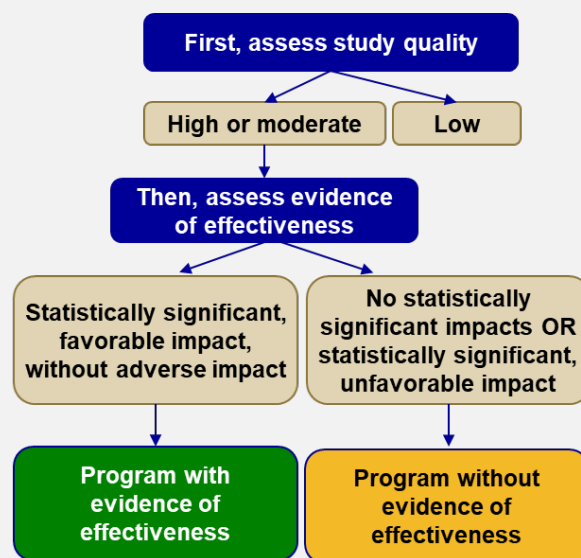
What is the TPP Evidence Review?

Since 2009, HHS has contracted with Mathematica to conduct the TPP Evidence Review.

This systematic review is a tool to help policymakers, practitioners, and other decision makers identify evidence-based TPP programs. At the federal level, HHS has used the findings in part to determine eligibility for federal grant funding for TPP programs. The review findings are also intended as a broader resource for states and local communities.

The TPP Evidence Review identifies and assesses studies of programs that aim to reduce teen pregnancies, STIs, and associated sexual risk behaviors. The review process is divided into two key stages (Figure 1). First, trained reviewers assess study quality and assign a quality rating to each eligible study to denote the quality and execution of the study's research design. This assessment accounts for five core elements of the study and yields a quality rating of high, moderate, or low. In the second stage, high- or moderate-quality studies are given a program effectiveness rating, which indicates whether the program or program component shows favorable effects on outcomes of interest. This brief describes the evidence standards, how those elements are used to assign study quality ratings, and the criteria used to assign a program effectiveness rating.

Figure 1. Stages for determining programs with evidence of effectiveness



For the TPP Evidence Review, both programs and program components are eligible if they intend to reduce rates of teen pregnancy, STIs, or associated sexual risk behaviors.

- **Programs** are typically some combination of educational, skill-building, and/or psychosocial intervention. Programs may be delivered either one-on-one to individuals or in groups, in any type of public, private, or institutional setting. Examples include classroom-based health curricula, individualized programs delivered by health professionals in clinics or other settings, community-based or after-school programs, and specialized programs for youth in the juvenile justice or child welfare systems.
- **Program components** must be (1) a clearly defined practice, procedure, policy, support, or organizational structure, potentially with documented steps for implementation with fidelity to facilitate replication; and (2) capable of being implemented independently, in conjunction with, or integrated into a TPP intervention. Examples of components that could be eligible for review include practices such as in-class condom demonstrations and text messaging as an enhancement to a well-defined TPP program.

What are the criteria of the study quality assessment?

When assessing the quality of a study, the TPP Evidence Review examines five elements of the study: (1) research design, (2) reassignment, (3) attrition, (4) baseline equivalence, and (5) confounding factors. Table 1 presents each of the five criteria and, for each, includes a brief definition and explanation for why it affects a study rating. In addition, the table provides some considerations for designing and implementing studies to improve the likelihood that a study meets evidence standards. The table also provides a link to additional resources, beyond the evidence review protocol, to learn more about these topics. Finally, the table also indicates whether the criterion applies to a randomized controlled trial (RCT) or a quasi-experimental design (QED).

Only studies that meet the eligibility criteria in the TPP Evidence Review protocol are reviewed against the evidence standards (Mathematica 2023). Some key eligibility criteria are:

- The study must have been conducted in the United States.
- The study must have completed data collection within the last 20 years.
- The majority of sample members must be 19 or younger.

Table 1. Criteria of the study quality assessment

Criteria	RCT	QED
Research design		
<p>Definition: Two types of designs are eligible for the TPP Evidence Review: (1) RCTs and (2) QEDs with external comparison groups.</p> <p>Why it affects the study rating: The research design affects how well differences in outcomes can be attributed to the intervention. RCTs are the stronger design for establishing a causal effect of the program. They ensure that intervention and comparison groups are equivalent on all measurable and unmeasurable characteristics; thus, they are the only study design eligible for the high quality rating.</p> <p>Considerations for studies: In RCTs, random assignment can be of individual youth to intervention or comparison conditions or of clusters (e.g., all youth in a set of classrooms are assigned to the intervention while all youth in another set of classrooms are assigned to the comparison condition).</p> <p>QEDs need to have an external comparison group to be eligible for a moderate rating; a pre-post design using only one group of (treated) youth cannot meet TPP Evidence Review standards.</p>	X	X
Reassignment		
<p>Definition: In RCTs, reassignment occurs when the randomly assigned units are not analyzed based on their initial assignment status.</p> <p>Why it affects the study rating: Moving participants from one study group to another, because of exposure, lack of exposure, or noncompliance, can produce bias in a study's impact estimate. For example, if low-motivation students assigned to the intervention group decide not to attend the program, and are analyzed as if they are in the comparison condition, the program will appear to be more effective than it truly is.</p> <p>Considerations for studies: Analyze all units according to their initial assignment status (i.e. conduct an intent-to-treat, or ITT, analysis). For instance, youth who do not attend a program should still be analyzed as part of the treatment group in an RCT. If units are reassigned to condition during the study, or if the impact analysis does not compare individuals based on their initially assigned condition, the study must demonstrate baseline equivalence and is not eligible for the highest rating.</p>	X	
Attrition		
<p>Definition: Attrition occurs when members of the originally randomly assigned sample do not have outcome data (e.g., they do not respond to the follow-up survey).</p> <p>Why it affects the study rating: The loss of participants can bias the impact estimates by creating differences in the observed and unobserved characteristics of the treatment and control groups. For example, if all of the sexually active youth in the intervention group drop out of the study, and the sexually active youth in the comparison group remain in the study, the program will appear to be effective at reducing sexual activity. However, this effect would be solely due to sample attrition.</p> <p>Considerations for studies: The evidence review factors in both overall attrition and the difference in the attrition rates between treatment and control groups when determining whether the risk of bias from attrition is so high that baseline equivalence must be established. Therefore, it is important to not only minimize overall attrition (i.e., maximize response rates across the full sample), but to also minimize differences in attrition rates across intervention and comparison groups (i.e., ensure that the response rates across conditions are similar). Non-consent after random assignment may be factored into attrition calculations. The TPP Evidence Review assesses attrition against thresholds established by the U.S. Department of Education's What Works Clearinghouse.</p> <p>Additional resource: "Sample Attrition in Teen Pregnancy Prevention Impact Evaluations." Research Brief. (Cole and Chizeck 2023).</p>	X	

Criteria	RCT	QED
Baseline equivalence		
<p>Definition: QEDs and RCTs with either reassignment or high attrition must show that the intervention and comparison groups are equivalent at baseline (pre-intervention) on age, race/ethnicity, and biological sex. Study authors might also statistically adjust for any baseline differences (see protocol for more details). For studies with sample members at least 14 years old at baseline, the study authors must also establish baseline equivalence on at least one outcome measure. Studies with younger sample members are exempt from establishing baseline equivalence of outcome measures because rates of sexual risk behaviors are typically low for this age group.</p> <p>Why it affects the study rating: Well-matched groups help to minimize the risk of bias from a designed, or sample loss or reassignment in an RCT. Similarly, if groups are very dissimilar at baseline on characteristics that influence outcomes, any post-intervention differences in outcomes may be due to these pre-existing differences at baseline, rather than being due to the intervention being studied.</p> <p>Considerations for studies: Collect extensive data at baseline on characteristics of sample participants that are expected to influence outcomes. At a minimum, this baseline data collection should include sample demographics and sexual behavior information (if age appropriate). Although not currently required by the review, baseline measures on knowledge, attitudes, and other personal characteristics may also help reduce the risk of bias.</p> <p>Additional resource: “Baseline Inequivalence and Matching.” Research Brief. (Cole and Agodini 2023)</p>	X	X
Confounding factors		
<p>Definition: A confounding factor of the research is not a part of the intervention but aligns with one of the study conditions.</p> <p>Why it affects the study rating: It is impossible to tell if differences in the outcomes are due to the intervention or to the confounding factor. For example, if a single school receiving the intervention is compared against a single comparison school not receiving the intervention, it may be the case that the two schools are systematically different in terms of the sexual risk profiles or other related characteristics of the youth that they serve.</p> <p>Considerations for studies: Make sure to have more than one group of participants in each study condition. For instance, it may be necessary to recruit additional schools for a QED or cluster RCT.</p> <p>Also, make sure there are no systematic differences in data collection between study conditions (e.g., data on one group is collected via survey and another group via administrative data).</p>	X	X

How does a study get a high or moderate quality rating?

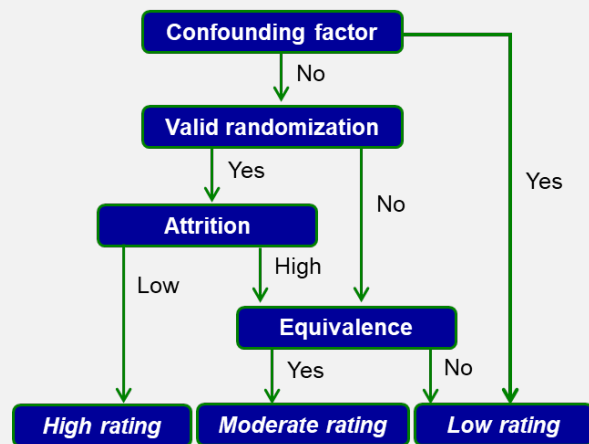
Figure 2 presents how each criteria factors into a study's quality rating. Only RCTs assessed as having low attrition, no reassignment, and no confounding factors can receive a high quality rating. QEDs and RCTs assessed as having high attrition or reassignment, that demonstrate baseline equivalence and have no confounding factors, can receive a moderate quality rating. Studies that do not meet either the high or the moderate standards receive a low rating.

How does a program or program component demonstrate evidence of effectiveness?

A program or program component that has at least one eligible study with a high or moderate quality rating is assessed for evidence of program effectiveness. If any of the studies rated high or moderate demonstrate statistically significant, favorable impacts on any outcomes of interest, and the analysis meets the criteria (see box below), the program or program component is noted as effective. Outcome domains of interest include sexual activity, contraceptive use, STIs, pregnancy, or birth.

The HHS TPP Evidence Review summarizes the research evidence across all available studies of the program or program component. Each study is reviewed individually, and then a rating is provided to characterize all evidence for a program or program component within each outcome domain. The review will be updated periodically to identify new evaluations and to improve upon the review criteria as best practices evolve.

Figure 2. Decision rules for the TPP Evidence Review quality ratings



Analytic criteria for the program effectiveness ratings

- Outcomes must be measured for either the full analytic sample or a subgroup defined by (1) biological sex or (2) sexual experience at baseline.
- Subgroups should not be defined by a post-random assignment characteristic (e.g., looking at a subset of those who had sex at the one year follow-up) because that analysis will produce a biased impact estimate.
- Statistical significance must be assessed with a two-tailed hypothesis test and a specified alpha level of $p < .05$
- Because only statistically significant findings contribute to ratings of program effectiveness, it is important to ensure sufficient study power when designing your evaluation.
- Intra-cluster correlation adjustments are required for cluster RCTs and QEDs.

References

Cole, R., and Agodini, R. (2023) "Baseline Inequivalence and Matching." Evaluation Technical Assistance Brief. Submitted to the Office of Population Affairs. Princeton, NJ: Mathematica.

Cole, R., and Chizeck, S. (2023). "Sample Attrition in Teen Pregnancy Prevention Impact Evaluations." Evaluation Technical Assistance Brief. Submitted to the Office of Population Affairs. Princeton, NJ: Mathematica.

Mathematica. (2023). "Identifying Programs That Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors Review Protocol Version 7.0."

This brief was written by Lauren Murphy and Jean Knab from Mathematica for the HHS Office of Population Affairs under contracts #HHSP233201300416G and HHSP233201500035I/75P00122F37068.

The TPP Evidence Review was not actively updated between 2018 and 2022.