Meta-Analysis of Federally Funded Teen Pregnancy Prevention Programs: Final Report

November 2019

Randall Juras and Meredith Kelsey
Abt Associates
Mark Lipsey and Katarzyna Steinka-Fry
Vanderbilt University
Jean Layzer
Belmont Research Associates

Submitted to:
Amy Farb
The Office of Population Affairs
Lisa Trivits and Sarah Oberlander
The Office of the Assistant Secretary for Planning and Evaluation
Tia Brown, Jessica Johnson, and Kenyatta Parker
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP233201500069I

Project Director: Randall Juras
Abt Associates
5001 S. Miami Blvd #210
Durham, NC 27703

Disclaimer: This publication was supported by Award No. HHSP233201500069I from the Office of Population Affairs (OPA). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of OPA or the U.S. Department of Health and Human Services.

Acknowledgements

We gratefully acknowledge financial support and technical guidance from the Office of Adolescent Health, the Office of the Assistant Secretary for Planning and Evaluation, and the Administration for Children and Families (ACF) within the U.S. Department of Health and Human Services. Contracting Office's Representative Amy Farb played a critical role in guiding the study and provided helpful comments on multiple drafts of this report. We also thank the following HHS staff for their feedback and other efforts on behalf of the study: Lisa Trivits, Sarah, Oberlander, Tia Brown, Jessica Johnson, and Kenyatta Parker.

Many faculty and staff at Abt Associates and Vanderbilt University contributed to this study. We would especially like to thank Emily Tanner-Smith at Vanderbilt University (now at the University of Oregon), who played an invaluable role leading the study's early design and analysis tasks. We would additionally like to acknowledge Alicia Hymel, Heather Hensman Kettrey, Jennifer Norvell, Leah Roberts, and Emily Farnen at Vanderbilt and Yeqin He at Abt Associates for their outstanding data coding work, and Austin Nichols at Abt Associates for his helpful comments on drafts of the article.

Contents

Executive Summary

Beginning in 2010, the U.S. Department of Health and Human Services (HHS) funded a large number of teen pregnancy prevention program grants through three grant programs: the Teen Pregnancy Prevention (TPP) Program, the Personal Responsibility Education Innovative Strategies (PREIS) program, and the Personal Responsibility Education Program (PREP). These grants were overseen by two agencies: the Office of Adolescent Health (OAH) within the Office of the Assistant Secretary for Health (OASH) and the Family and Youth Services Bureau (FYSB) within the Administration for Children and Families (ACF). These agencies also oversaw many rigorous independent evaluations of the grant-funded programs.

Evaluation Rigor

*HHS-funded evaluations were typically designed to adhere to the **HHS Teen Pregnancy Prevention (TPP) Evidence Review** standards for research quality. That protocol specifies minimum standards for study design, attrition, baseline equivalence, assignment to research groups, and confounding factors.*

Before this effort, evidence of effectiveness for most teen pregnancy prevention program models came from single studies, often conducted by the program developer many years earlier.[1] This increase in federal investment in evaluation reflected recognition of the need for independent testing to assess the replicability of evidence-based programs in contemporary contexts and with diverse populations, as well as to generate evidence to support or refute those original study findings. Beyond replications, HHS funded several new and untested programs, and sponsored evaluations of almost all of them. The explicit goal of these efforts was to expand and strengthen the research base, so that future decision makers would have a more robust body of evidence upon which to base their selection among program models.

In 2015, as one of several subsequent steps in a carefully sequenced research agenda, OAH, ACF, and HHS's Office of the Assistant Secretary for Planning and Evaluation (ASPE) funded a quantitative synthesis of the findings from all of these evaluation efforts. Abt Associates Inc., in partnership with the Peabody Research Institute of Vanderbilt University and Belmont Research Associates, was selected to conduct the analysis. This is the meta-analysis final report.

What was the purpose of the meta-analysis?

The evaluations funded by HHS beginning in 2010 resulted in a set of findings on the effect of individual program models on behavioral outcomes. The Abt team intended to use the data generated by these evaluations to ask, what, if any, common elements could be identified as enhancing program effectiveness? Though evidence about the general effectiveness of program models is essential, evidence on what strategies or program components work better and for which youth populations could help program developers design more effective programs and help practitioners select appropriate ones to implement. Equipped with this information, they could make a better match among programs, their own resources, the characteristics of their communities, and their local youth populations. The synthesis project represents an initial attempt to address these and related questions.

---

[1] Evaluations conducted by program developers can be rigorous, and their impact findings are usually modest. However, implementation of the model is likely to be much more carefully controlled than in replication studies, and independent evaluations often reveal smaller impact estimates (Borman et al. 2003; Petrosino and Soydan 2005). This reality makes replication and independent testing more crucial.

Which evaluations were included?

Studies were eligible for the meta-analysis if they used rigorous research methods and reported program effects on youth sexual behavior, pregnancy, or sexually transmitted infections. In particular, the study must have assessed program effects using experimental or high-quality quasi-experimental research designs, comparing youth who received a teen pregnancy prevention program with youth who did not receive similar services. The study must have also had a sample of at least 10 study participants whose average age was between 10 and 19 years old. HHS allowed us to review all studies, regardless of favorability or statistical significance, thus the presence or absence of significant findings was not a criterion for inclusion.

We screened 66 reports of independent teen pregnancy prevention studies funded by OAH and ACF and completed before April 15, 2019. We determined that 61 of these reports were eligible for review. Some of them included multiple studies; that is, their authors evaluated the effectiveness of a program model using more than one study sample of youth or evaluated two or more program models using separate samples of youth. All told, a total of 53 separate studies representing 57,364 youth and 45 distinct program models met the inclusion criteria and were analyzed.

To enhance the meta-analysis, we also analyzed participant-level data on 48,635 youth from 34 of the studies for which such data were available.

How was the meta-analysis conducted?

The Abt team extracted data from each of the 53 studies using standard systematic reviewing and meta-analysis procedures.[2] In order to compare results across studies, we represented program effects on risky sexual behaviors and consequences as statistical *effect sizes*. We used a meta-regression framework to examine whether these effect sizes were associated with aspects of program design, program implementation, or the demographic composition of the samples. We used a similar approach to analyze factors associated with attendance and retention.[3] To examine whether individual-level participant characteristics such as age, gender, and race/ethnicity were associated with behavioral effect sizes, we used an approach that combined participant-level data from the 34 studies with aggregate data from the other 19 studies for which participant-level data were not available. We categorized findings with $p<.05$ as significant and findings with $p<.10$ as marginally significant.

Findings

On average, the programs in the meta-analysis had small favorable effects on their most important outcomes (that is, those that the program evaluators pre-specified as confirmatory). In particular, there was a small, statistically significant overall effect as well as marginally significant effects on two particular outcomes: *recent pregnancy* and *ever had sex*.

On average, as reported in the studies we analyzed, effects were small and effect sizes did not vary much across studies. As a result, we had little success identifying the potential sources (e.g., program or population characteristics) of what limited variation there was. However, we did identify one program characteristic potentially linked to program effectiveness that may be of interest to program developers

---

2 For details, see Borenstein et al. (2009).

3 *Retention* is defined as the average proportion of participants attending 75 percent or more of the program sessions across all program periods.

and practitioners: program setting. In particular, **programs delivered in classrooms had less favorable effects** than did programs delivered in other settings such as clinics, after-school settings, participants' homes, and online.

We did not observe statistically significant associations between effect sizes and other individual program design features or program implementation characteristics. We found no evidence that effect sizes are associated with any of the individual participant characteristics we examined such as gender, race/ethnicity, and age.

We did, however, identify two program characteristics that may be worth further examination in future studies: **Programs designed exclusively for girls, and programs that delivered services to individual youth showed particular promise.** Although the differences in program effects between these programs and others were not statistically significant at conventional levels (p=.12 and p=.13), the average effects for such programs were particularly favorable.

Finally, we found no evidence that evaluation methods affected the estimated effect sizes.

Our review found that most programs were implemented with fidelity to the intended program model, but that attendance and retention rates varied considerably across studies. This variation was correlated with some observable program elements—for example, both attendance and retention rates were significantly higher in programs that offered condom demonstrations. However, because it is difficult to separate these elements from other factors, we caution against over-interpreting these findings.

Conclusions

On average, these HHS-funded programs caused a small reduction in risky sexual behaviors and their consequences. This is promising news for the field. The finding that programs delivered in classrooms are less effective than programs delivered in other settings—and hints that programs designed for girls and programs with individualized delivery may be more effective than other programs—may also offer some direction for program developers, implementers, and evaluators.

1. Introduction

Beginning in 2010, the U.S. Department of Health and Human Services (HHS) funded a large number of teen pregnancy prevention program grants through three grant programs: the Teen Pregnancy Prevention (TPP) Program, the Personal Responsibility Education Innovative Strategies (PREIS) program, and the Personal Responsibility Education Program (PREP). These grants were overseen by two agencies: the Office of Adolescent Health (OAH) within the Office of the Assistant Secretary for Health (OASH) and the Family and Youth Services Bureau (FYSB) within the Administration for Children and Families (ACF).

As part of the grant funding, each agency required rigorous independent evaluations assessing the program effects of a subset of the interventions. Evaluations were a mixture of grantee-led and federal studies, almost all of which estimated the overall effect of a specific program on adolescents' risky sexual behaviors and consequences.[4] To ensure grantees delivered high-quality studies, they received intensive evaluation technical assistance (TA) and training from a federal evaluation contractor. In addition, an inter-agency workgroup developed a common core of survey items to be used in the federal evaluations (and a subset of behavioral outcome measures to be used in all evaluations), a step that resulted in consistently measured and reported outcomes.

Finally, HHS was committed to releasing the evaluation findings regardless of their favorability or statistical significance. This commitment to evaluation rigor, consistency in outcome measures, and transparency presents a remarkable opportunity to look across a large number of studies to learn more about what program or contextual elements contribute to program effectiveness.

1.1 Analysis of Findings

The analysis reported on here applies **meta-analytic techniques** to systematically investigate the findings generated by the HHS evaluations. Meta-analysis is a tool that can be used to search for patterns in the evidence in a principled and systematic manner—for example, to determine whether certain types of program models tend to produce effects in certain outcome domains. In particular, this meta-analysis:

- Estimates the *overall effect* of the HHS-funded programs on key behavioral outcomes;
- Addresses questions about whether elements of *program design*, *program implementation*, and *participant characteristics* are associated with the estimated program effects on behavioral outcomes.
- Assesses the degree to which similar elements are associated with program attendance and retention rates, regardless of the ultimate effect of the program on behavioral outcomes.
- Examines the extent to which variation in evaluation design and methodology is related to the resulting estimates of effects.

[4] All OAH TPP Tier 2 grantees, all FYSB PREIS grantees, and a subset of OAH TPP Tier 1 grantees (those receiving annual grants in excess of $1 million) were required to contract with an independent evaluator to assess the impact of the intervention. At the same time, HHS directly funded and monitored its own federal evaluation contracts, the designs for which incorporated a small number of TPP Tier 1 and Tier 2 grantees, PREIS grantees, and PREP grantees.

1.2 Purposes

The findings are intended to serve three related purposes: (1) to help program developers design more effective programs, either new programs or informed adaptations; (2) to help practitioners select and implement programs most appropriate to the characteristics of their communities and their local youth populations; and (3) to help guide funding decisions by federal, state, and local entities.

1.3 Reporting

This is the final report of the meta-analysis. Results from an interim analysis of short-term findings from 44 studies was previously published (Juras et al. 2017). This current updated analysis incorporates findings from an additional nine studies and longer-term follow-ups from 15 studies.

2. Context

This chapter provides background information on the teen pregnancy prevention programs that were funded by HHS, and the studies that were analyzed for this report.

2.1 HHS's Teen Pregnancy Prevention Programs

This meta-analysis examines evaluations of pregnancy prevention programs funded by HHS through the TPP Program; the PREIS program; and PREP state, Tribal, and competitive grants. In this section, we provide a brief overview of each of these three initiatives.

2.1.1 TPP Program Grants

The Teen Pregnancy Prevention Program is administered by OAH within HHS. It is one of six tiered evidence-based initiatives across the federal government in which a majority of program funding supports the implementation of programs with some prior evidence of effectiveness.[5] The first cohort of grants, funded in 2010, supported competitively awarded projects in two categories: Tier 1 grants (75 percent of grant funds) supported the replication of evidence-based program models; and Tier 2 grants (25 percent of grant funds) supported the implementation of new and untested teen pregnancy prevention programs.

Tier 1 Grants

With the approximately $75 million dollars available in 2010 for Tier 1 ("replication") grants, OAH awarded 75 five-year grants for implementation of one or more of 28 program models initially identified as having evidence of effectiveness in preventing teen pregnancy, sexually transmitted infections (STIs), and/or adolescent risky sexual behaviors. These 28 program models had been identified through the *HHS Teen Pregnancy Prevention (TPP) Evidence Review*, a comprehensive systematic review of evaluations of teen pregnancy prevention programs.

The diverse grantees included school districts, community-based organizations, faith-based organizations, state and city departments of health, research organizations, and institutes of higher education. Ultimately, the 75 Tier 1 grantees implemented a total of 23 of the 28 program models identified by the *HHS TPP Evidence Review* in 2010. A subset of these grantees were required to conduct independent evaluations of the funded program models (see Section 2.2 below).

Tier 2 Grants

In 2010, approximately $25 million was available for Tier 2 ("innovative") grants to implement a program model that was new, a significant adaptation of an evidence-based program model, or a previously developed but untested program model. Tier 2 funds supported 19 five-year grants in 14 states

[5] The six initiatives are the Teen Pregnancy Prevention and Home Visiting Programs administered by HHS; the Investing in Innovation Fund (i3) administered by the U.S. Department of Education; the Social Innovation Fund administered by the Corporation for National and Community Service; and the Trade Adjustment Assistance Community College and Career Training (TAACCT) program and the Workforce Innovation Fund, both administered by the U.S. Departments of Labor and Education. For the TPP Program specifically, what constituted evidence was having "at least one favorable, statistically significant program effect on at least one sexual risk behavior or reproductive health outcome of interest (sexual activity, number of sexual partners, contraceptive use, STIs, or pregnancy)" from a study that meets established criteria for the quality and execution of the research design (Lugo-Gil et al. 2016).

and the District of Columbia; 18 of these grantees conducted independent evaluations (see Section 2.2 below).

2.1.2 PREP Grants

Administered by FYSB, the Personal Responsibility Education Program (PREP) awards state, Tribal, and competitive grants to educate youth on both abstinence and contraception. These three grant programs are intended to target youth in areas with high teen birth rates and specific high-risk populations such as homeless youth, youth in foster care or juvenile justice systems, youth living in rural areas, youth from minority groups, and pregnant or parenting youth. Similar to Tier 1 TPP Program grantees, PREP grantees implemented effective, evidence-based program models previously identified through a comprehensive systematic review, or they substantially incorporated elements of these program models.

The State PREP program was authorized in 2010 as part of the 2010 Patient Protection and Affordable Care Act. The majority of funds ($55.25 million of $75 million annually) was to support formula grants to states and territories. States can administer their project directly or through sub-awards to public or private entities. In the first two years of available funding, 49 states and territories applied and received funding. For the Tribal PREP program, guidelines are similar to those of the State PREP program, but funds are available specifically to Tribes and Tribal communities to develop and implement projects for American Indian/Alaska Native (AI/AN) youth. The first cohort of 15 grantees received funding for 5-year projects in 2011.

The competitive PREP program supports grants to local organizations. Beginning in 2013, organizations and entities in states and territories that had not applied for State PREP funds were eligible to submit competitive applications for funding. Thirty seven grants were awarded to organizations in FY2013 .

2.1.3 PREIS Grants

Also administered by FYSB as part of the Adolescent Pregnancy Prevention Program, the Personal Responsibility Education Innovative Strategies (PREIS) grants target high-risk or vulnerable youth ages 10–19 who are aging out of foster care, homeless, pregnant or parenting, or live in areas with high teen birth rates. Similar to Tier 2 TPP Program grantees, PREIS grantees implement innovative or untested strategies with a focus on interventions that fill gaps with new, promising program models.

In 2010, $10 million was available to fund competitive grants. FYSB awarded 11 grants. Grantees were required to collect performance measure data on participant engagement and fidelity of implementation, as well as to conduct rigorous impact evaluations designed to meet the standards of the *HHS TPP Evidence Review*. (See Section 2.2 below.)

2.2 Evaluations of HHS Teen Pregnancy Prevention Programs

HHS funded a large number of rigorous evaluations of TPP, PREP, and PREIS grant-funded programs in an effort to expand the evidence base on effectiveness. This report synthesizes results from five types of evaluations:

1. Grantee-led evaluations conducted by TPP Program grantees through cooperative agreements with OAH;
2. Grantee-led evaluations conducted by PREIS Program grantees through cooperative agreements with FYSB;
3. OAH/ASPE-led TPP Replication Study;

4. OAH-led Adolescent Pregnancy Prevention Approaches (PPA) Study; and
5. ACF-led Personal Responsibility Education Program (PREP) Multi-Component Evaluation.

Both of the OAH-led evaluations and the PREP evaluation were conducted under contract by research firms. Together, these five sets of evaluations represent a large proportion—more than half—of all recent studies of the effectiveness of teen pregnancy prevention program models in the United States.

2.2.1 TPP Program Grantee-Led Evaluations

To generate more evidence on what works to prevent teen pregnancy, certain TPP Program Tier 1 grantees and all Tier 2 grantees were required to conduct evaluations of their interventions.[6]

By definition, all Tier 1 program models had been previously evaluated and had demonstrated effects on at least one sexual health outcome. At the time grants were awarded, all but 1 of the 28 program models meeting the *HSS TPP Evidence Review* standards had demonstrated evidence of effectiveness through a single study, so that the generalizability of the findings to other populations or settings was largely unknown.[7]

OAH believed that more independent evaluations were needed to better understand the robustness of the original findings and to provide information to policymakers and program implementers on the effectiveness of the programs in contemporary conditions and with broader populations. As a result, OAH required all Tier 1 grantees that had been awarded at least $1 million per year in funding to conduct rigorous impact evaluations. Evaluators were required to be independent of the funded entity. Together, Tier 1 grantee studies produced evidence about the replicability and generalizability of eight of the 28 previously evaluated program models. Tier 2 grantee-led evaluations contributed evidence on the effectiveness of 17 previously untested program models.

Each grantee-led evaluation was conducted through a cooperative agreement with OAH. To ensure that evaluations would generate the best possible evidence, OAH provided extensive high-quality technical assistance to grantees through a federal evaluation contractor. The technical assistance effort is described in detail by Knab, Cole, and Zief (2016) and Zief, Knab, and Cole (2016).

OAH also took steps to ensure consistency across evaluations so that results could be more easily compared. Study teams were required to collect and report data on a commonly defined set of behavioral outcome measures. Through the technical assistance provider, evaluators were encouraged to employ a common set of analysis methods, which were pre-specified in design reports and subject to multiple rounds of independent review.

Findings from the grantee-led evaluations were released on a rolling basis in 2015 and 2016.

---

[6] The 75 Tier 1 grantees each applied for funding in one of four ranges. The 16 awardees in the top two ranges (Range C: $1 million–1.5 million/year; Range D: $1.5 million–$4 million/year) were expected to conduct a rigorous independent evaluation. Grantees could spend up to 25 percent of program funds per year on evaluation (or a maximum of $500,000 per year). Some evaluations of Tier 1 and Tier 2 programs were conducted as part of the TPP Replication Study and the PPA Study.

[7] At the time of the original list, *Be Proud Be Responsible* was the only identified program on the eligible list that had more than one study demonstrating effectiveness (Jemmott et al. 1999; Jemmott 1992).

2.2.2 PREIS Grantee-Led Evaluations

Similar to Tier 2 TPP Program grantees, PREIS grantees were each expected to conduct independent evaluations of their projects. The grantee-led evaluations received the same evaluation technical assistance that was provided to the TPP Program grantee-led evaluations. Also like the TPP Program grantee-led evaluations, PREIS evaluations were designed to meet the *HHS TPP Evidence Review* standards. Evaluators were expected to carefully document the intervention for possible replication by others, conduct process and outcome evaluations, and disseminate findings. Reports from seven PREIS evaluations were released between 2016 and 2018, in time to be considered for this meta-analysis. Three of the grantee-led PREIS evaluations were included in the PPA Study described below.

2.2.3 TPP Replication Study

The TPP Replication Study was conducted by Abt Associates.[8] Using a series of rigorous randomized experiments, the study tested replications of three widely used evidence-based (Tier 1) program models being implemented by TPP Program grantees (*¡Cuídate!*, *Reducing the Risk*, and the *Safer Sex Intervention*), each in three sites, to determine their effectiveness when implemented with fidelity across different settings and populations. Short-term impact reports for each of the three program models were released in 2016; long-term impact reports were released in 2018.

2.2.4 PPA Study

The Adolescent Pregnancy Prevention Approaches (PPA) Study, conducted by Mathematica Policy Research, rigorously tested the effectiveness of untested and innovative programs in six federally funded sites—three OAH TPP Program Tier 2 grantees and three FYSB PREIS grantees—and diverse settings across the United States. Because of the diversity of settings, target populations, and program strategies, each site represents a test of the effectiveness of a single program model. Interim reports from all six evaluations and final reports from five evaluations were released in time to be considered for this meta-analysis.

2.2.5 PREP

The Personal Responsibility Education Program (PREP) Multi-Component Evaluation was an effort to assess the effects of programs funded through the state and competitive PREP grant programs, and to document how those programs were implemented in the field. The evaluation was conducted by Mathematica Policy Research and used a random assignment design in four specific PREP-funded sites. HHS made findings from three of these four PREP studies available to our team in time to be considered for this meta-analysis.

2.3 Summarizing Findings from the Evaluations

The large number of studies sponsored by HHS beginning in 2010 represents a dramatic (and purposeful) increase in the evidence base for teen pregnancy prevention programs. To guard against publication bias in this rapidly expanding literature, HHS committed from the outset to releasing research findings

[8] Three of the authors of this meta-analysis (Randall Juras, Meredith Kelsey, and Jean Layzer) were co-authors on the TPP Replication Study reports. To avoid a conflict of interest, these authors were recused in the meta-analysis from all determinations of eligibility, full-text coding, and quality for those studies. Such determinations instead were made by staff at Vanderbilt University with no input from Juras, Kelsey, or Layzer.

regardless of favorability or statistical significance. Final evaluation reports from all of the TPP Program grantee-led evaluations are held in a collection at the National Library of Medicine and are available on the OAH website.[9] Final reports from the TPP Replication Study and PPA Study are also available on the OAH website. Evaluation reports from PREP are available on the ACF website and reports from PREIS will be made available on the ACF website.[10,11]

Results from 21 of the TPP Program grantee evaluations were published in a September 2016 supplemental issue of the *American Journal of Public Health* (*AJPH)* titled "Building the Evidence to Prevent Adolescent Pregnancy."[12] The *AJPH* supplemental issue included a synthesis of the impact findings that summarized data and themes, assessed implementation quality, and counted statistically significant behavioral outcomes (Farb and Margolis 2016). The synthesis found that 21 percent of the studies of programs in Tier 1 and 36 percent of the studies in Tier 2 demonstrated some evidence of effectiveness (positive and statistically significant findings for at least one key behavioral outcome).

A primary goal of this meta-analysis is to rigorously and systematically search for patterns in the findings across the five types of evaluations—for example, to determine whether certain types of program models tend to produce effects in certain outcome domains. The purpose is twofold: that program developers can design more effective programs, and practitioners can select the programs that are most appropriate for their communities and populations. By pooling across many studies, the meta-analysis might overcome some of the key limitations that prevent each of the underlying studies from answering these questions. The primary limitation for any one study is the lack of variability in key program or participant characteristics that is inherent in studies of a single program model.

The number of evaluations funded by HHS to date is sufficient to provide reasonably good estimates of the overall average effects on the primary outcomes. However, there are not enough studies to support good estimates of differences between subgroups of studies. As a result, we were unable to assess the association between complex combinations of program and demographic characteristics and behavioral outcomes. Therefore, the report's findings are limited to associations between key characteristics, considered individually, and behavioral outcomes.

---

9 https://www.hhs.gov/ash/oah/evaluation-and-research/grantee-led-evaluation/index.html

10 https://www.acf.hhs.gov/opre/research/project/personal-responsibility-education-program-prep-multi-component

11 https://www.acf.hhs.gov/fysb/programs/adolescent-pregnancy-prevention/evaluation/program-impacts

12 http://ajph.aphapublications.org/toc/ajph/106/S1

3. Research Design

This chapter sets forth this report's research questions, methods, and analytic strategy.

3.1 Research Questions

The meta-analysis's major research questions, developed in collaboration with HHS and outside experts in the field, drive the investigation of whether and how aspects of program design and implementation or characteristics of participants are associated with the programs' effects on risky sexual behaviors in adolescents.

The answers could help program developers refine and target their interventions more effectively. They also could help agency staff and grantees select program models that reflect the needs and characteristics of the youth they plan to serve.

As is standard in meta-analysis, we examine the overall effectiveness of the whole group of HHS-funded programs, pooling across evaluations that report those programs' effects on conceptually similar outcomes (e.g., effects related to recent unprotected sexual activity) and also pooling across all evaluations and all effect sizes.

In addition to examining overall effectiveness, the meta-analysis has three primary research questions exploring variation in effect sizes:

1. **Which, if any, aspects of program design are associated with program effectiveness?** Program design characteristics that could potentially influence effectiveness include, for example, program focus, size and composition of participant groups, and program duration. (The full list of program design moderators is shown in Section 3.3.2.) In addition, there could be a difference in effectiveness between evidence-based programs (Tier 1) and new and innovative programs (Tier 2).
2. **Which, if any, aspects of program implementation are associated with program effectiveness?** Examples here include fidelity of implementation, participant attendance, and retention.
3. **Which, if any, characteristics of participants are associated with program effectiveness?** Participant characteristics that could potentially influence effectiveness include gender, race/ethnicity, age, and prior sexual experience.

The meta-analysis also addresses two secondary research questions:

4. **To what extent is variation in study methods related to the resulting estimates of program effectiveness?** Attributes of interest include, for example, study type (i.e., randomized experiment or quasi-experiment), rates of attrition, and type of comparison group (active vs. passive).[13]

[13] An *active* comparison group is one in which members of the control group receive alternate services as part of the evaluation, such as driving skills training. A *passive* comparison group receives no services from the evaluation, only assessments, but may receive other "business-as-usual" services that are not associated with the evaluation. Because OAH's funding was focused on geographic areas with high risk and few resources, the business-as-usual condition did not typically involve substantial pregnancy prevention services.

5. **Which, if any, program characteristics are associated with <u>better attendance and retention</u> of participants?** To address this question, all the aspects of program design and implementation or participant characteristics noted above (except attrition) are considered as potential predictors of attendance and retention.

The reason for considering these two research questions as secondary is that their findings have limited applicability and/or may be difficult to interpret. Research question (RQ) 4 could be of interest to researchers as well as agency staff and policymakers across a range of topic areas who might be called on to approve or specify research designs, but RQ 4 is unlikely to be of interest to program developers or implementers. RQ 5 could be of interest to program funders and implementers, but the findings might be difficult to interpret or apply (e.g., if program elements found to be associated with higher attendance/retention were to conflict with program elements that are associated with program effectiveness).

3.2 Data Sources

The meta-analysis collected no new data. Instead, it used existing data from two sources: findings from HHS-funded evaluations and grantee-provided performance measure data.

3.2.1 Evaluation Findings

We identified and selected studies for the meta-analysis using a rigorous screening process, which is described in Appendix A. In total, 61 reports of HHS-funded evaluations, describing the results of 53 independent studies of 56 intervention programs, were determined to be eligible.[14] (A full list of references to eligible studies is provided in Appendix B. A list of studies determined to be ineligible, along with reasons for ineligibility, is provided in Appendix C.) Most grantees collected and reported data from three time points: a baseline prior to program delivery, a short-term follow-up, and a longer-term follow-up.[15] Some grantees collected data at fewer or additional follow-up intervals. In addition to descriptive data from the study, we used the evaluation reports to code outcomes related to each available time point.

To enhance the meta-analysis, several authors of eligible studies provided the data on individual study participants that they used in their analyses. Through OAH, we requested this information from the authors of reports received prior to October 2016. In total, report authors provided such individual participant data (IPD) on 48,635 youth from 34 studies.

3.2.2 Performance Measure Data

Every HHS-funded program grantee was required to collect a standard set of performance measurement data that included fidelity, attendance, and retention metrics. These data were reported to OAH every six

---

[14] In some studies, two or more interventions were evaluated using a shared control group. We classified each of these as a single study ($k = 1$) with more than one intervention group ($g > 1$). There were two such studies in the sample. First, Fronius (2016) evaluated three interventions: (1) *Salud y Éxito*, (2) *Más Que un Sueño*, and *(3) ¡Cuídate!* that shared one control group. Second, Tanner (2016) evaluated two focal interventions, (1) *Possessing Your Power (PYP)* and (2) *Choosing the Best (CTB)*, that were contrasted with *Discovery*.

[15] In some cases, short- and longer-term findings from the same study were presented in separate reports. Less commonly, a single report presented results from more than one study (e.g., each of the TPP Replication Study reports presented results from three studies).

months. HHS made these data available to the Abt team for the purpose of coding attendance and retention outcomes. We obtained corresponding performance data for the three PREP evaluations from the PREP implementation reports.

3.3 Definitions of Outcomes and Moderators

Each of the research questions specified above concerns the association between an outcome of interest, such as *recent sexual activity* or *better attendance*, and one or more moderators. A **moderator** is a characteristic of a program or study participant that potentially could be associated with program outcomes; that is, we might observe that programs with certain characteristics have systematically larger effect sizes than do programs without those characteristics.

Outcomes and potential moderators were specified, defined, and coded in preparation for the analysis as described below. The specific outcomes and moderators selected for analysis were drawn from the evaluation reports we reviewed and the IPD we received. Because no primary data were collected for the meta-analysis, only outcomes and moderators that were consistently available from one or more of these two sources could be analyzed.

3.3.1 Outcome Measures

To address the meta-analysis's research questions, we coded two types of outcomes: **behavior and consequence measures**, and **participation measures**. All evaluations reported on at least one outcome of each type. We coded program effects as standardized effect sizes, with higher, positive values favoring the treatment group.[16]

RQs 1–4: Behavior and Consequence Measures

The primary outcomes of interest for the first four research questions of this meta-analysis were risky sexual behaviors and consequences. All HHS-funded programs were intended ultimately to reduce one or more risky sexual behaviors and thus reduce the rate of unintended pregnancies. The specific behaviors targeted varied by program model (e.g., some program models encouraged teens to delay engaging in sexual activity, whereas other program models sought to increase condom and other birth control use in sexually active teens). As a result, study authors reported various outcome measures, which we assigned to nine categories as defined below:[17]

1. *Ever had sex* (defined as vaginal, oral, or anal sex)
2. *Recent sexual activity* (defined as vaginal, oral, or anal sex during any recall period since the baseline survey was administered; usually reported for a 90-day period)
3. *Recent unprotected sexual activity* (defined as vaginal sex without a condom or other birth control during a recall period after baseline; usually reported for a 90-day period)
4. *Number of sexual partners* (for vaginal, oral, or anal sex; over any time period)

---

[16] Effect size estimates were coded as either odds ratios (for binary measures) or Hedges' *g* small-sample corrected standardized mean difference effect sizes (for continuous measures). All analyses were conducted using the log odds ratio (with Hedges' *g* effect sizes converted to log odds ratios for this purpose), also known as the logit. A technical discussion of how effect sizes and standard errors were calculated is provided in Section 1.4 of the report's technical supplement.

[17] *Baseline* is defined as at the time of enrollment in the study.

5. *Number of sexual experiences* (for vaginal, oral, or anal sex; over any time period)
6. *Proportion of sexual experiences that were unprotected* (defined as vaginal, oral, or anal sex without a condom or other birth control; over any time period)
7. *Sexually transmitted infections* (defined as any lifetime or recent measure of STI incidence)
8. *Ever pregnant* (during lifetime; for boys, this outcome is defined as having ever got someone pregnant)
9. *Recent pregnancy* (during any recall period since baseline; for boys, this outcome is defined as having recently got someone pregnant)

All 53 studies reported on an outcome in at least one of these nine categories, but no single outcome was uniformly reported across all studies.

The authors of each study pre-specified, in their analysis plans and prior to reviewing data, a small number of outcomes (typically one or two) as "confirmatory." Authors had chosen these outcomes as the best indicators of whether each program was on track to achieve, or had achieved, its intended goals.[18] We identified which outcomes had been specified as confirmatory in each evaluation. Because study authors put considerable thought into specifying confirmatory outcomes that aligned with the programs' logic models, we followed their lead and focused on these confirmatory outcomes throughout the meta-analysis.[19] To the degree that evaluators chose appropriate outcomes to report (i.e., outcomes corresponding to the programs' goals), this approach captures the best measure of behavioral "success" for each program model.[20]

RQ 5: Participation Measures

The meta-analysis's fifth research question explores the extent to which participant attendance and retention were affected by program characteristics. To address this research question, we coded two continuously measured outcomes:

1. *Participant attendance* (defined as the average percentage attendance rate across all program sessions)

---

[18] The purpose of specifying a confirmatory outcome is to guard against the heightened possibility of spurious findings that arises when simultaneously testing multiple hypotheses. The probability of at least one spurious finding among *n* independent tests is $1-(1-\alpha)^n$. With α equal to five percent (i.e., a 95 percent confidence interval), the probability of a spurious finding is five percent for one test, 23 percent for five tests, 40 percent for 10 tests, and so on. As a result, the OAH evaluation technical assistance contractor required grant-funded evaluations to specify confirmatory outcomes. The other HHS-funded evaluations also adopted this approach.

[19] Section 3.7 of the technical supplement presents the results of identical analyses that incorporate all reported effect sizes for any of the nine outcomes from each study. We reference these findings in the main text where relevant.

[20] For example, consider an evaluation of a program that reported program effects on a single confirmatory outcome of *recent sexual activity* and an evaluation of a program that reported program effects on two confirmatory outcomes of *recent unprotected sexual activity* and *number of sexual partners*. Our analysis would incorporate all three outcomes while statistically adjusting for the fact that two of them were reported in the same evaluation report. Our analysis can compare the program effects across different outcomes because all effects were coded as effect sizes using a common unit of measurement.

2. *Participant retention* (defined as the average proportion of participants attending 75 percent or more of the program sessions across all program periods)

We obtained data on these two outcomes, which were self-reported by grantees and likely of varying quality, directly from HHS. These data were available for subsets of 51 studies (attendance) and 50 studies (retention). We also used both measures as moderators in the analysis of behavioral outcomes, as described below.

3.3.2 Moderators

We coded effect size moderators in a number of categories, corresponding to the focus of each research question—program characteristics (design and implementation), participant characteristics, and study methods.[21]

RQ 1: Program Design Moderators

To investigate the extent to which program design (as characterized in the evaluation reports) is associated with the measured effect of teen pregnancy prevention programs, we examined seven categories of related moderators:

1. *Program focus* (defined as the philosophy or theory of change that guides program activities: depending on its philosophy or driving principles, a program might prioritize some messages or purposefully omit others)
2. *Program components* (defined as the types of activities that participants engage in during the course of the program; for example, role plays, condom demonstrations, service learning, and parent activities)
3. *Group size* (defined as the number of participants in a typical group)
4. *Group composition* (defined as whether lessons are delivered to same- or mixed-gender groups. For example, a classroom-based program could deliver the curriculum to boys and girls together in mixed-gender classrooms, or might instead separate youth into boys-only and girls-only groups)
5. *Gender specificity (girls only)* (defined as whether the program was designed for girls only. For example, the *Safer Sex Intervention* was designed specifically for girls who visit health clinics; boys are not eligible. In contrast, *Reducing the Risk* is intended for both boys and girls, although they may be separated into single-gender groups for program delivery)
6. *Program length* (defined as the amount of time youth spend in the program; includes the cumulative amount of time youth spend in the program—the total amount of calendar time that elapses from the beginning to the end of the program and the number of hours actually spent receiving services—as well as how often program staff meet with youth)

[21] Section 1.5 of the report's technical supplement provides additional detail on how moderators were defined and coded for analysis. Specific instructions on how moderators were coded from reports are provided in the coding manual, which is reproduced in its entirety as Section 1.2.2 of the technical supplement. In most cases, we relied on each report's characterization of program elements, rather than consult the original program manuals, in case any modifications had been made to the original designs.

7. *Level of prior evidence* (defined based on tiered funding structure; Tier 1 programs are evidence-based (TPP Tier 1 program grants; state, Tribal, and competitive PREP program grants) whereas Tier 2 programs are new or untested (TPP Tier 2; PREIS program grants).

RQ 2: Program Implementation Moderators

To measure the extent to which program implementation (as characterized in the evaluation reports) is associated with the measured effect of teen pregnancy prevention programs, we examined three categories of related program moderators:

1. *Program setting* (defined as the type of location in which the program was typically delivered; e.g., classroom, community)
2. *Program delivery personnel* (defined as the type of staff generally charged with delivering the program to youth)
3. *Implementation characteristics* (defined as the strength with which the program was implemented—in terms of both fidelity to the original program model and attendance and retention of participants)

RQ3: Participant Characteristics Moderators

To measure the extent to which participant characteristics are associated with the measured effect of teen pregnancy prevention programs, we examined four related moderators:

1. *Gender* (defined as the proportion of the sample who identify as boys)
2. *Race/ethnicity* (defined as the racial and ethnic composition of the sample)
3. *Age* (defined as the average age of youth in the sample at baseline)
4. *Risk level* (defined as the proportion of the control/comparison group that reported having ever been sexually active—including vaginal, oral, or anal sex—at follow-up)[22]

Participant characteristics were coded at the study level for all eligible studies. For studies that provided IPD, we also coded each individual participant's gender, race/ethnicity, and age.

RQ 4: Study Methods Moderators

To measure the extent to which each study's design, methods, and procedures are associated with the measured effect of teen pregnancy prevention programs, we examined five related moderators:

1. *Study design* (defined as the study's expected level of internal validity based on whether or not it was a randomized experiment)

---

[22] Because only studies that used rigorous research methods were eligible for the meta-analysis, we know that the treatment and control/comparison groups in each study have similar characteristics. Thus, the control group's behavior at follow-up should be a close proxy for the behavior that would have been observed in the treatment group if the group had not been exposed to the intervention. We had originally intended to use *sexual experience at baseline* as a crude proxy for risk level, but we had to drop that measure because it was not available for many studies. After developing the protocol, we determined that sexual behavior among control group participants at follow-up (rather than at baseline) was a better measure of each sample's risk level, because it assesses sexual risk at a slightly older age, when youth are more likely to engage in such behaviors. This and other deviations from the study's pre-specified protocol are described in Section 2.2 of the report's technical supplement.

2. *Overall attrition* (defined as the proportion of study participants who were lost to follow-up)
3. *Differential attrition* (defined as the degree to which different proportions of study participants were lost to follow-up in the treatment groups and control/comparison groups, which is a potential indicator of bias)
4. *Active comparison condition* (defined as control group receipt of an intervention not related to teen pregnancy as part of the evaluation—e.g., nutrition information or driving skills—rather than exposure to only a business-as-usual condition)
5. *Study rated as inconclusive* (defined as a study having received a rating of "inconclusive" in a recent synthesis published in the *American Journal of Public Health* (Farb and Margolis 2016): "Evaluations placed in this category experienced challenges with either program implementation, quality of the evaluation, or both")

RQ 5: Attendance and Retention

The meta-analysis's fifth research question explores the association between attendance/retention and the moderators for program design, program implementation, and participant characteristics defined for RQs 1–3.

3.4 Analytic Approach

We used two types of data from the HHS-funded evaluations—aggregate data and IPD—to answer the meta-analysis's research questions. Aggregate data were used in analysis of all five research questions. IPD from a subset of 34 studies were used in analysis of only RQ 3 (e.g., Are teen pregnancy prevention programs more or less effective for girls?).

Our basic strategy for analyzing both types of data was to use a regression-based framework, which allowed us to examine the effect of a single moderator while adjusting for other moderators.[23] To ensure the most rigorous and unbiased analysis, we pre-specified all analytic methods before conducting any data analysis (Abt Associates 2016).[24]

For the meta-regression, we set the *p*-value threshold for significance at a conservative value of five percent for individual results; that is, we considered results significant (and discuss them in the report's text) only if the 95 percent confidence interval did not include zero effect. We refer to results with $p < .10$ as "marginally significant." When simultaneously testing a group of conceptually similar moderators, we also assessed the overall (omnibus) significance of the group as a whole and used this information to help inform our discussion.

These procedures help guard against spurious findings from the large number of hypothesis tests, but they also increase the chance that some true associations might be overlooked. As a result, this report presents two kinds of results:

- *Main findings*, which are findings in which we have a high level of confidence; and

[23] A more comprehensive discussion of the study's analytic approach is provided in Section 2.1 of the report's technical supplement.

[24] Deviations from the protocol are provided in Section 2.2 of the technical supplement.

- *Suggestive findings*, which are findings that we considered intriguing (e.g., because they help to explain the main findings) but in which we do not have a high level of confidence.

Suggestive findings are useful for hypothesis generation but should not be taken as definitive evidence of effects.

4. Characteristics of Studies Included in the Meta-Analysis

This chapter describes some basic characteristics of the studies and programs that were included in the meta-analysis sample. These sample descriptive characteristics can be categorized into three broad categories: (1) study methods, (2) program characteristics, and (3) participant characteristics.

4.1 Methodological Characteristics of the Studies

The methodological quality of the studies was high. Almost half of effect sizes were measured more than 12 months after program completion.

Most studies (89 percent) randomly assigned participants to conditions, the average overall attrition rate at first follow-up was 25 percent (*SD* = 18 percent), and the average differential attrition between intervention and comparison groups was three percent (*SD* = four percent). A small number of studies suffered from high attrition and differential attrition, with a maximum attrition rate of 88 percent at the final follow-up for one study.

Slightly more than a third of studies (36 percent) used an active counterfactual condition, where students received an alternative program that provided information on a topic unrelated to sexual health such as obesity or driving skills. Appendix D provides brief summaries of the intervention and comparison conditions evaluated in each study. Exhibit 4-1 below describes the sample of studies used to evaluate these programs.

Effect sizes for confirmatory outcomes were reported at varying durations of follow-up, with many studies reporting program effects at multiple follow-up intervals.[25] Among confirmatory outcomes only, 12 percent of effect sizes were measured up to three months after the end of the program, 20 percent were measured between three and six months after the program, 18 percent were measured between six and nine months after the program, five percent were measured between nine and 12 months after the program, and 45 percent were measured at a longer-term follow-up (more than 12 months after program completion).[26]

---

[25] The meta-regression models used for analysis are capable of incorporating multiple effect sizes per study, as well as effect sizes measured at different follow-up intervals. The timing of outcome measurement was included as a covariate in the regression model. We found no evidence that follow-up interval was associated with effect size.

[26] In total, including both confirmatory and other outcomes, 387 effect sizes were obtained. Of all reported effect sizes, 12 percent were measured up to three months after the end of the program, 20 percent were measured between three and six months after the program, 27 percent were measured between six and nine months after the program, three percent were measured between nine and 12 months after the program, and 38 percent were measured at a longer-term follow-up (more than 12 months after program completion).

Exhibit 4-1: Characteristics of the Included Studies (*k* = 53)

	All Studies (*k* = 53)		Studies Providing IPD (*k* = 34)		
	n (%) or Mean (SD)	Valid *N*	*n* (%) or Mean (SD)	Valid *N*	Range
Study Methods and Quality					
Design					
Randomized controlled trial (individual)	25 (47%)	53	16 (47%)	34	
Randomized controlled trial (cluster)	22 (42%)	53	15 (44%)	34	
Controlled quasi-experimental design	6 (11%)	53	3 (9%)	34	
Active comparison condition	19 (36%)	53	12 (35%)	34	
Study rated as inconclusive[a]	10 (19%)	53	7 (21%)	34	
Attrition					
Attrition at first follow-up	.25 (0.18)	51	.25 (0.17)	34	.03–.62
Attrition at last follow-up	.32 (0.19)	48	.31 (0.19)	17	.03–.88
Differential attrition[b]	.03 (0.04)	113	.02 (0.02)	164	.00–.14
Post-Test Assessment Timing[b]					
0 < X ≤ 3 months	14 (12%)	121	8 (10%)	84	
3 < X ≤ 6 months	24 (20%)	121	19 (23%)	84	
6 < X ≤ 9 months	22 (18%)	121	9 (11%)	84	
9 < X ≤ 12 months	6 (5%)	121	2 (2%)	84	
12 < x months	55 (45%)	121	46 (55%)	84	

Notes. *k* = number of studies. IPD = individual participant data. SD = standard deviation. Means and standard deviations shown for continuous measures; frequencies and percentages shown for dichotomous measures.

[a] See Farb and Margolis (2016).

[b] Estimates calculated at effect size level (*N* = 121).

4.2 Characteristics of the Programs Studied

The majority of evaluated teen pregnancy prevention programs focused on adolescent sexual health and were delivered to large groups of youth. About half were implemented in classroom settings, with the most frequently implemented components being role plays, condom demonstrations, and parent activities. Almost 40 percent of the programs were evidence-based, selected from the list of programs compiled by the* HHS TPP Evidence Review *in 2010.

Approximately 38 percent of the 56 evaluated programs appeared on the list of evidence-based programs compiled by the *HHS TPP Evidence Review* in 2010. Most programs (89 percent) had either a sexual health (64 percent) or a youth development focus (25 percent). The most frequently implemented program components were role plays (used in 61 percent of programs), condom demonstrations (30 percent), parent activities (23 percent), games (21 percent), and reflective exercises (20 percent).

Programs were typically delivered to large groups of more than 10 youth (61 percent) and in classroom settings (46 percent), with groups comprising both girls and boys (73 percent). Ten programs were intentionally designed for girls only.[27] The personnel delivering the programs varied widely, including

[27] Because only one program, Wise Guys, was designed specifically for boys it was not possible to analyze boys-only programs as a separate group. The evaluation of Wise Guys was grouped with mixed-gender programs for the meta-analysis.

health educators (43 percent) and classroom teachers (13 percent). Most programs were delivered at least once a week, and the median program had between 10 and 30 total contact hours with each participant over the course of the program. Only 18 percent of programs had a duration of more than one year.

Overall, teen pregnancy prevention programs were implemented with high fidelity to the key concepts of their program models. The strength of implementation was more variable; some programs struggled to retain participants and ensure adequate exposure to the intervention.

Program implementation, attendance, and retention data indicated that most programs were delivered with high fidelity, although some struggled with attendance and retention. Facilitators reported that on average, 95 percent (range 69–100 percent) of all teen pregnancy prevention program sessions adhered to the specified program model; the average attendance rate across all project periods was 81 percent (range 35–100 percent); and the average rate of participant retention (defined as the average proportion of participants attending 75 percent or more of program sessions) was 76 percent (range 20–100 percent).

Exhibit 4-2 presents descriptive statistics for the key features of the 56 teen pregnancy prevention programs evaluated in the 53 eligible studies.

Exhibit 4-2: Characteristics of the Interventions/Programs (*k* = 53; *g* = 56)

Intervention Characteristic	*n* (%) or Mean (SD)	Range (if applicable)
Primary Program Focus		
Abstinence only	2 (4%)	
Sexual health	36 (64%)	
Youth development	14 (25%)	
HIV/AIDS prevention	3 (5%)	
Reproductive health services	1 (2%)	
Program Component		
Condom demonstrations	17 (30%)	
Service learning	6 (11%)	
Role plays	34 (61%)	
Games	12 (21%)	
Reflective exercises	11 (20%)	
Mentoring/tutoring	1 (2%)	
Individualized counseling	2 (4%)	
Direct provision of health services	5 (9%)	
Parent activities	13 (23%)	
Community outreach	1 (2%)	
Positive role model	7 (13%)	
Group Size		
Individualized	6 (11%)	
Small groups (<10)	10 (18%)	
Large groups	34 (61%)	
Online	3 (5%)	
Other (mixed individual/group)	3 (5%)	
Program Setting		
Classroom	26(46%)	
Community	14 (25%)	

Intervention Characteristic	*n* (%) or Mean (SD)	Range (if applicable)
Other setting	16 (28%)	
Group Composition		
Youth separated into single-gender groups	14 (27%)	
Program delivered in mixed-gender groups	38 (73%)	
Gender Specificity (girls only)		
Program designed for girls only[a]	10 (18%)	
Program designed for both genders or for boys	46 (82%)	
Program Delivery Personnel		
Medical professionals	1 (2%)	
Health educators	24 (43%)	
Classroom teachers	7 (13%)	
Peer educators	2 (4%)	
Mixed	8 (14%)	
Other[c]	14 (25%)	
Program Length		
Frequency of Contact (Valid *k* = 50)		
Daily	4 (8%)	
3–4 times per week	5 (10%)	
1–2 times per week	33 (66%)	
Less than weekly	8 (16%)	
Duration (Valid *k* = 51)	36.72 (47.88)	1–189
< 4 weeks	11 (22%)	
4–11.9 weeks	13 (26%)	
12–51.9 weeks	18 (35%)	
52+ weeks	9 (18%)	
Contact Hours (Valid *k* = 55)	52.52 (162.76)	1–1134
< 10 hours	19 (35%)	
10–29.9 hours	25 (45%)	
30–49.9 hours	6 (11%)	
50+ hours	5 (9%)	
Implementation Characteristic		
Fidelity (Valid *k* = 43)	.95 (0.06)	.69–1
Attendance (Valid *k* = 51)	.81 (0.16)	.35–1
Retention (Valid *k* = 50)	.76 (0.20)	.20–1
Evidence-based program	21 (38%)	

Notes. *k* = number of studies, g = number of intervention groups. IPD = individual participant data. SD = standard deviation. Means and standard deviations shown for continuous measures; frequencies and percentages shown for dichotomous measures. Valid *k* = 53 and *g* = 56, unless noted otherwise.

[a] Gender-specific (girls only) programs are designed to be delivered exclusively to girls. Because program content is tailored for girls, it would be inappropriate to deliver the program to boys. The single boys-only program evaluation in the meta-analysis sample was grouped with mixed gender programs for all analyses, including in this table.

[c] The "other" category includes online delivery, delivery by CD, and delivery by facilitators or program staff that were recruited based on their specific background and skills and then trained in program delivery.

4.3 Characteristics of the Program Participants

Across the studies, 40 percent of participants were boys, and the gender composition of the individual studies varied widely. More than one third of the participants were Black, about the same fraction were Hispanic, and almost all of the remainder were White.

The demographic composition of the total sample varied considerably across studies. On average, 40 percent of study participants were boys; there were several all-girl samples and one all-boy sample. Study samples were mixed in terms of race/ethnicity (mean percentage Black = 36 percent, Hispanic = 40 percent, White = 25 percent). The average age was 14.5 years (SD = 2.16) at baseline.

The sample was riskier in its behavior than the national average, but not exceptionally high-risk: approximately 36 percent of control group members were sexually active at follow-up, compared with approximately 25 percent of similar-age youth nationwide.

Only half of the studies reported participants' baseline sexual activity. Among those studies that did, on average, 30 percent of youth reported *recent sexual activity* at baseline (i.e., at an average age of 14.5 years), and 13 percent of youth reported *recent unprotected sexual activity* at baseline.

Most studies reported sexual activity at one or more follow-up intervals, so we also used post-test sexual activity in the control/comparison group as a proxy for the riskiness of each sample. On average, 36 percent of control/comparison group youth were sexually active at the first follow-up—the timing of which varied considerably across studies, but with a majority of follow-ups more than six months after program completion. In comparison, approximately 25 percent of 10th-graders nationwide (which roughly corresponds to the sample's average age at first follow-up) report recent sexual activity (Kann et al. 2018). Exhibit 4-3 presents descriptive statistics for the key features of the study participants at baseline.

Exhibit 4-3: Characteristics of Youth Participants (*k* = 53; *g* = 56)

Participant Characteristic	All Studies (*n* = 56)		Studies Providing IPD (*n* = 34)		
	n (%) or Mean (SD)	Valid *N*	*n* (% or Mean (SD)	Valid *N*	Range
Percentage boys	.40 (0.21)	56	.37 (0.20)	34	0–.75
Percentage Black	.36 (0.33)	50	.36 (0.31)	32	.01–1
Percentage Hispanic	.40 (0.31)	49	.35 (0.27)	32	.03–1
Percentage White	.25 (0.21)	45	.25 (0.22)	29	0–.83
Average age	14.49 (2.16)	52	14.46 (2.18)	34	11–19
Proportion *recent sexual activity* at baseline	.30 (0.30)	28	.34 (0.33)	20	0–1
Proportion *recent unprotected sexual activity* at baseline	.13 (0.18)	39	.16 (0.22)	23	0–.74
Proportion control group sexually active at post-test	.36 (0.30)	52	.41 (0.31)	34	0–1

Notes. *k* = number of studies; *g* = number of intervention groups. IPD = individual participant data. SD = standard deviation. Means and standard deviations shown for continuous measures; frequencies and percentages shown for dichotomous measures.

5. Overall Effects of the Evaluated Programs

This chapter summarizes the findings from the 53 studies that contributed aggregate data, for each of the nine ***behavior and consequence outcomes*** reported in the studies we reviewed, seven of which appeared as confirmatory (most important) outcomes in at least one study. These results describe the effect of all of the programs, on average.

How to Read Tables in This Chapter

Exhibit 5-1displays effects for confirmatory outcomes; the last row of Exhibit 5-1 summarizes findings across all confirmatory outcomes. Exhibit 5-2 displays effects for all outcomes reported.

*The first column in each exhibit shows the **number of studies** that contributed data for each outcome.*

*The second column shows the **total number of effect sizes reported**, which is always as large or larger than the number of studies contributing data for that outcome. For example, a single study that reported program effects for vaginal sex in the past 90 days and oral sex in the past 90 days as its two confirmatory outcomes would contribute two effect sizes in the "recent sexual activity" outcome category.*

*The last two columns in each exhibit summarize the **findings for each outcome**—first showing the effect size as a log odds ratio (for binary outcomes) or Hedges' g effect size (for numeric outcomes) and then the 95 percent confidence interval.*

On average, the HHS-funded programs in the meta-analysis sample had favorable and significant effects on their confirmatory outcomes. In particular, there was a small, statistically significant (p = .04) average effect across all outcomes across all studies. There also were marginally significant effects on "recent pregnancy" (p = .05) and "ever had sex" (p = .07).

Exhibit 5-1 below shows the average overall effects of HHS-funded programs on their confirmatory outcomes. *Confirmatory* outcomes, of which there were typically one or two per study, are those that program evaluators selected prior to data analysis as the most important indicators of whether the program was on track to achieve its long-term goals or had already achieved them. All told, evaluators selected confirmatory measures in seven of the nine behavior and consequence outcome categories. The most commonly reported confirmatory outcome was recent unprotected sexual activity, which appeared as a confirmatory outcome in 32 studies.

For the seven confirmatory outcomes in Exhibit 5-1, the overall treatment effect size on average across all studies is expressed as a log odds ratio. **A positive number indicates an effect favoring the treatment group (i.e., the intervention was effective) and a negative number indicates an effect favoring the control group (the intervention was not effective)**. For these seven outcomes, we found both a favorable and significant average effect across all outcomes (p = .04). We also found favorable and marginally significant effects on two specific outcomes: *ever had sex* (p = .07) and *recent pregnancy* (p = .05).

Generally speaking, the overall effect sizes presented in Exhibit 5-1 are small. For the outcome *ever had sex* (log odds ratio 0.07), the average effect size expressed in percentage points is 1.1 percentage points—assuming a base rate of 19 percent, which is the average across studies that reported program effects for

this outcome. This means that if the percentage of youth ever sexually active in the control group at follow-up were 19.0 percent, then the estimated average percentage of youth sexually active in the treatment group would be 17.9 percent. Other outcomes have similarly small effects.

One exception is the outcome *recent pregnancy* (log odds ratio 0.26), for which the overall effect size is somewhat larger. Expressed in percentage points, the average effect is 3.8 percentage points—now assuming a base rate of 16 percent, which is again the control group average for studies reporting program effects for this outcome. This means that if the percentage of youth recently pregnant in the control group were 16.0 percent, then the estimated average percentage of youth recently pregnant in the treatment group would be 12.2 percent.

Log Odds Ratio

*A **log odds ratio** is the most appropriate measure for aggregating effects for binary variables across studies, but is not a metric that lends itself to easy interpretation.*

In this section of the report, we roughly illustrate the magnitude of key estimates by translating the log odds ratio into percentage point units for some given base rate (i.e., some given proportion of youth engaging in the activity in the absence of the intervention).

Exhibit 5-1: Overall Effects of Teen Pregnancy Prevention Programs for Confirmatory Outcomes

Outcome Construct	# of Studies	# of Effect Sizes Reported	Effect Size Expressed as Log Odds Ratio	
			Log Odds Ratio	[95% Confidence Interval]
Ever had sex	22	26	0.07†	[–0.01, 0.14]
Recent sexual activity	17	26	–0.05	[–0.18, 0.08]
Recent unprotected sexual activity	32	48	0.05	[–0.04, 0.15]
Proportion of recent sexual experiences that were unprotected	1	1	–0.29	[–0.85, 0.27]
Ever pregnant	4	4	0.19	[–0.68, 1.06]
Recent pregnancy	12	12	0.26†	[0.00, 0.52]
Number of sexual partners[a]	2	2	0.08	[–1.27, 1.44]
Average effect for all outcomes	52	119	0.07*	[0.00, 0.14]

Notes. Effects were coded such that positive log odds ratios or Hedges' *g* effect sizes indicate effects favoring the treatment group (e.g., less sexual activity, less pregnancy). This analysis omits one eligible study for which we did not have sufficient data for estimating confirmatory effect sizes (Lauby et al. 2017).

[a] Hedges' *g* effect size converted to LOR, as detailed in the technical supplement.

* $p < .05$, † $p < .10$.

Across all reported outcomes (including those not labeled as confirmatory), there was a small and marginally significant (p = .09) average effect. For specific outcomes, program effects were favorable but mostly small and statistically insignificant. One exception was for "recent pregnancy": on average, the 19 studies reporting this outcome had a statistically significant effect.

Exhibit 5-2 summarizes the findings from all 53 studies that contributed aggregate data, this time for each of the 385 effect sizes (in nine ***behavior and consequence*** categories) that were coded from study reports, regardless of whether those effects were labeled as confirmatory outcomes. Outcomes in seven of the nine

categories were measured on binary (yes/no) scales (e.g., *recent sexual activity*) whereas two of the nine were coded on numeric scales (e.g., *number of sexual partners*). All effects were coded such that **a positive number indicates an effect favoring the treatment group (i.e., an effective intervention)**.

For these outcomes, we found both a favorable and marginally significant average effect across all outcomes ($p = .09$) and a favorable and statistically significant effect on *recent pregnancy* ($p < .05$). Again, generally speaking, the overall effect sizes presented in Exhibit 5-2 are small with the exception of *recent pregnancy*, which has an effect size similar to the effect on *recent pregnancy* in Exhibit 5-1.

Exhibit 5-2: Overall Effects of Teen Pregnancy Prevention Programs for All Outcomes

Outcome Construct	# of Studies	# of Effect Sizes Reported	Effect Size Expressed as Log Odds Ratio	
			Log Odds Ratio	[95% Confidence Interval]
Ever had sex	29	56	0.04	[–0.03, 0.11]
Recent sexual activity	27	92	0.02	[–0.05, 0.09]
Recent unprotected sexual activity	41	146	0.05	[–0.02, 0.12]
Proportion of recent sexual experiences that were unprotected	1	1	–0.29	[–0.85, 0.27]
Sexually transmitted infections	11	11	0.17	[–0.35, 0.70]
Ever pregnant	8	16	0.12	[–0.20, 0.44]
Recent pregnancy	19	24	0.24*	[0.04, 0.45]
Number of sexual partners[a]	5	10	0.03	[–0.04, 0.09]
Number of sexual experiences[a]	1	1	0.03	[–0.12, 0.18]
Average effect for all outcomes	53	385	0.05[†]	[–0.01, 0.11]

Notes. Effects were coded such that positive log odds ratios or Hedges' *g* effect sizes indicate effects favoring the treatment group (e.g., less sexual activity, less pregnancy).

[a] Hedges' *g* effect size converted to LOR, as detailed in the technical supplement.

* $p < .05$, [†] $p < .10$.

One striking finding in Exhibit 5-2 is that although there was an average program effect on *recent pregnancy*, there was not an overall effect on any of the behavioral outcomes thought to be precursors to pregnancy (such as *recent sexual activity* or *recent unprotected sexual activity*). A potential explanation for this apparent paradox lies in the fact that only 19 of the 53 studies reported an effect size in the *recent pregnancy* category, whereas many more studies contributed effect sizes for other behavioral outcomes. Restricting the analysis to only those 19 studies that reported on *recent pregnancy*, we found a significant favorable effect on *recent unprotected sexual activity*—meaning that in these 19 studies, there was both an average effect on pregnancy and an effect on one of its precursor behaviors.[28]

[28] Average effects of these 19 programs on all measured behavioral outcomes, and for the 34 programs that did not measure recent pregnancy, are shown in section 3.9 of the technical supplement. For the outcome of *recent unprotected sexual activity*, the effect size was $b = 0.10$, with $p < .05$.

Effects of teen pregnancy prevention programs on behavior and consequence outcomes were similar across studies.

In addition to calculating the average effect size for each outcome, we assessed the degree to which the estimated treatment effects varied across studies, based on statistical measures of between-study variability and the proportion of total variability in effects due to between-study variability.[29] We found that for all seven confirmatory behavior and consequence outcomes, the program effects were remarkably homogeneous across studies.

It is clear that there was some variation in program effects: some of the studies found statistically significant evidence of program effects on at least one behavioral outcome, whereas others did not. Nonetheless, as a whole, the reported program effects did not substantially vary across studies.

[29] These measures are reported in Section 3.1 of the technical supplement.

6. Results for Primary Research Questions

This chapter presents the study's main findings for the three primary research questions: that is, whether any individual program design feature, program implementation characteristic, or characteristic of participants is associated with measured program effectiveness. The discussion of results in this chapter is limited to findings that are supported by a high level of statistical confidence. Results that we considered intriguing but in which we have less confidence are discussed in Chapter 8 on suggestive findings.

How to Interpret Regression Coefficients in This Chapter's Tables

Each table shows regression coefficients within blocks of similar moderators. Regression coefficients are shown in ***column "b"*** *in each exhibit.*

The regression coefficient indicates the magnitude and direction of the effect associated with each individual moderator.

In ***moderator blocks showing mutually exclusive categories*** *(e.g., primary program focus), the sign of the regression coefficient indicates whether programs of that type were more or less effective, on average, than programs in the reference category (indicated by "Ref."). In* ***all other moderator blocks****, the sign of the regression coefficient indicates whether programs with a given attribute (e.g., condom demonstrations) were more or less effective, on average, than programs without that attribute.*

In all cases, positive coefficients indicate (more) favorable effects.

6.1 Which, If Any, Aspects of Program Design Are Associated with Program Effectiveness?

No single aspect of program design was significantly associated with teen pregnancy prevention program effectiveness.

Using the aggregate data collected from the final evaluation reports, we fitted a series of meta-regression models to examine whether any of the seven teen pregnancy prevention *program design* features (moderators) described in Section 3.2.2 was associated with the observed effects on youth sexual behavior. In these models, data on the dependent (outcome) variable comprised all reported effect sizes for any of the study's seven confirmatory behavior and consequence outcomes . Effect sizes were coded so that higher, positive values favor the treatment group.

The model results summarized in Exhibit 6-1 below indicate that no single characteristic of program design was significantly associated with the magnitude of the effect size for confirmatory outcomes at the pre-specified threshold for statistical significance, $p < .05$.. (Results with larger p-values are discussed in Chapter 8.)

In particular, there were no significant differences in effect sizes across programs categorized according to *level of prior evidence* (program tier), with Tier 1 (evidence-based) and Tier 2 (new or untested) programs generating similar effect sizes.

Nor were there significant differences across programs in confirmatory effect sizes categorized according to whether the primary *program focus* was sexual health, youth development, or some other focus. None of the eight *program components* we tested (i.e., condom demonstrations, service learning, role plays,

games, reflective exercises, direct provision of health services, parent activities, and positive role model) was significantly associated with program effects on confirmatory outcomes.

Across all programs, the characteristics of *group composition* (provision of services in same-gender groups or mixed-gender groups), *group size*, and *gender specificity* (programs designed for girls only) were not significantly associated with average effect sizes at a high level of statistical confidence. There were no significant associations between effect sizes and *program length* (having at least one weekly contact with service recipients or number of contact hours).

Exhibit 6-1: Relationships between Program Design Features and Average Effect Sizes for Confirmatory Outcomes

	b	95% Confidence Interval
Level of Prior Evidence (Program Tier)		
Tier 2 program	Ref.	
Tier 1 program	–0.10	[–0.24, 0.03]
Intercept	0.11*	[0.03, 0.20]
	$F = 2.65$, $p = .11$	
Program Focus		
Sexual health	Ref.	
Youth development	0.03	[–0.14, 0.21]
Other	0.24	[–0.40, 0.88]
Intercept	0.04	[–0.02, 0.10]
	$F = 0.51$, $p = .62$	
Program Components		
Condom demonstrations	0.09	[–0.06, 0.23]
Service learning	0.15	[–0.31, 0.61]
Role plays	–0.05	[–0.22, 0.12]
Games	0.10	[–0.13, 0.32]
Reflective exercises	0.10	[–0.08, 0.27]
Direct provision of health services	0.38	[–0.27, 1.03]
Parent activities	–0.01	[–0.19, 0.16]
Positive role model	–0.14	[–0.35, 0.07]
Intercept	0.03	[–0.07, 0.12]
	$F = 1.61$, $p = .27$	
Group Size		
Individualized	Ref.	
Small groups (<10)	–0.26	[–0.53, 0.01]
Large groups	–0.25	[–0.52, 0.02]
Other (mixed individual/group)	–0.27	[–0.56, 0.01]
Intercept	0.29	[0.01, 0.57]
	$F = 1.55$, $p = .29$	

	b	95% Confidence Interval
Group Composition		
Program delivered in mixed-gender groups	Ref.	
Youth separated into single-gender groups	0.08	[–0.09, 0.25]
Intercept	0.05	[–0.02, 0.12]
	$F = 1.02$, $p = .33$	
Gender Specificity (girls only)		
Program designed for girls only	0.16	[–0.05, 0.37]
Intercept	0.04	[–0.03, 0.10]
	$F = 2.87$, $p = .12$	
Program Length		
At least weekly contact	–0.15	[–0.34, 0.03]
Contact hours	0.00	[–0.00, 0.00]
Intercept	0.18	[–0.00, 0.36]
	$F = 1.17$, $p = .43$	

Notes. *b* = unstandardized meta-regression coefficients, *F* = omnibus *F*-statistic for meta-regression model, Ref. = reference category. All meta-regression models estimated using robust variance estimation to handle statistically dependent effect sizes. The analytic sample size was *n* = 52 studies and 119 effect sizes for all analyses.

* $p < .05$.

6.2 Which, If Any, Aspects of Program Implementation Are Associated with Program Effectiveness?

Programs implemented in classrooms had less favorable effects than programs implemented in other settings (e.g., clinics, after-school, participants' homes, online). No other aspect of program implementation captured in the reports was significantly associated with teen pregnancy prevention program effectiveness.

The analyses described in this section used the same confirmatory outcome data as the models described in the previous section, but here the explanatory variables included in the models were measures of program implementation. The results summary table (Exhibit 6-2 below) can be read in the same way as Exhibit 6-1 above.

We found a significant relationship between *program setting* and program effects. In particular, we categorized programs according to whether the program setting was a classroom, community center, or some other setting.[30] Programs implemented in classrooms produced significantly less favorable effects than programs implemented in other settings ($p < .05$).

[30] The "Other" category comprises 16 evaluations of programs that were administered in settings other than schools or community centers: clinics (five evaluations), after-school (four evaluations) online or computer self-administration (three evaluations), participants' homes (three evaluations), and foster care agencies (one evaluation). Six of the 16 evaluations were implemented in mixed settings that included some classroom-based or community-center-based delivery alongside one of the other settings.

There were no significant differences in effect sizes across programs categorized according to whether the *program delivery personnel* were health educators, classroom teachers, or other staff. *Implementation characteristics* of implementation fidelity, mean attendance rates, and mean retention rates were not associated with average effect sizes (Exhibit 6-2).

Exhibit 6-2: Relationships between Program Implementation Features and Average Effect Sizes for Confirmatory Outcomes

	b	95% Confidence Interval
Program Setting		
Other	Ref.	
Community	–0.06	[–0.25, 0.13]
Classroom	–0.17*	[–.33, –0.01]
Intercept	0.17*	[0.03, 0.31]
	$F = 2.85$, $p = .09$	
Program Delivery Personnel		
Classroom teachers	Ref.	
Health educators	0.02	[–0.16, 0.20]
Other	0.05	[–0.15, 0.25]
Intercept	0.04	[–0.12, 0.21]
	$F = 0.17$, $p = .85$	
Implementation Characteristics		
Fidelity	–0.10	[–1.78, 1.58]
Mean attendance	3.42	[–0.97, 7.82]
Mean retention	–2.04	[–5.00, 0.92]
Intercept	–1.09	[–3.26, 1.07]
	$F = 1.28$, $p = .33$	

Notes. *b* = unstandardized meta-regression coefficients, *F* = omnibus *F*-statistic for meta-regression model, Ref. = reference category. All meta-regression models estimated using robust variance estimation to handle statistically dependent effect sizes. The analytic sample size for implementation characteristics (fidelity, attendance, retention) was n=42 studies and 101 effect sizes.

* $p < .05$.

6.3 Which, If Any, Characteristics of Participants Are Associated with Program Effectiveness?

No single characteristic of participants we tested was associated with program effectiveness.

The analyses conducted to address the question of whether characteristics of participants were associated with program effectiveness used two different data sets. The first analysis used the same aggregate data used in the models described in Sections 6.1 and 6.2. The results, shown in Exhibit 6-3, revealed no significant associations between program effectiveness and the percentage of study participants who were boys, Black, or Hispanic, their average age, or the base rate of participant risk (i.e., the proportion of participants in the comparison group who reported ever having sex at the first post-test assessment, used

as a proxy for the risk level that would have been expected in the treatment group absent the intervention).[31,]

Exhibit 6-3: Relationships between Participant Characteristics and Average Effect Sizes for Confirmatory Outcomes

	b	95% Confidence Interval
Participant Characteristics		
Percentage boys	–0.28	[–0.84,0.28]
Percentage Black	–0.01	[–0.44,0.42]
Percentage Hispanic	–0.02	[–0.43,0.39]
Average age	–0.03	[–0.14,0.09]
Risk (control event rate)	0.24	[–0.50,0.98]
Intercept	0.48	[–0.86,1.81]
	$F = 0.67$, $p = .66$	

Notes. *b* = unstandardized meta-regression coefficients, *F* = omnibus *F*-statistic for meta-regression model, All meta-regression models estimated using robust variance estimation to handle statistically dependent effect sizes.

The second set of analyses used the IPD provided by grantees. Using IPD, we were able to examine program effects for subgroups of participants defined by gender, race/ethnicity, and age. (In contrast, Exhibit 6-3 above shows the relationship between treatment effects and the percentage of each sample who were in each subgroup). Individual participant data were consistently available for four of the nine ***behavior and consequence*** outcomes: *ever had sex*, *recent sexual activity*, *recent unprotected sexual activity*, and *ever pregnant* (i.e., pregnancy for girls, causing pregnancy for boys).

We conducted the analysis separately by outcome, and we found no evidence that teen pregnancy prevention program effects on any of these four outcomes were significantly different from zero for any of the participant subgroups examined.[32,]

[31] Exhibit 6-3 shows the results from a single multivariable regression model, meaning that the estimated effect for each characteristic adjusts for all other characteristics in the model. Bivariate correlations between each characteristic and effect size are shown in Section 3.6 of the report's technical supplement.

[32] Exhibits showing the estimated program effect for each participant characteristic, with 95% confidence intervals, are provided in Section 3.2 of the report's technical supplement.

7. Results for Secondary Research Questions

This chapter presents the study's main findings for the two secondary research questions: that is, whether variation in study methods is related to estimates of program effectiveness; and whether any program characteristics are associated with better attendance and retention of program participants. As in Chapter 6, the discussion of results in this chapter is limited to findings that are supported by a high level of statistical confidence. Results that we considered intriguing but in which we have less confidence are discussed in Chapter 8 on suggestive findings.

7.1 Are Study Methods Related to the Resulting Estimates of Program Effectiveness?

Neither the rigor of study design nor the strength of study implementation was associated with the size of estimates of teen pregnancy prevention program effectiveness.

Using the same outcome data used in the models described in Sections 6.1, 6.2, and the first set of analyses in Section 6.3, we tested whether any of five study features was related to the magnitude of effect size estimates. As Exhibit 7-1 shows, none of these features—randomized experiments versus quasi-experiments, attrition rates of study participants, differential attrition (a measure of the difference between treatment and control group attrition rates at each follow-up), use of an active comparison group (e.g., nutrition information or driving skills, as opposed to business as usual), and a study rating of "inconclusive" in Farb and Margolis's (2016) synthesis—had a significant association with the magnitude of effect sizes. Collectively, these five study features did not explain a significant amount of outcome variance (*F*-test *p*-value = .60).[33]

Exhibit 7-1: Relationships between Study Methods and Average Effect Sizes

	b	95% Confidence Interval
Study Method		
Randomized controlled trial	0.16	[–0.33, 0.66]
Overall attrition	0.21	[–0.23, 0.65]
Differential attrition	0.42	[–2.73, 3.58]
Active control group	–0.07	[–0.24, 0.11]
Study rated inconclusive[a]	–0.21	[–0.44, 0.02]
Intercept	–0.09	[–0.68, 0.49]
	F = 0.77, *p* = .60	

Notes. *b* = unstandardized meta-regression coefficients, *F* = omnibus *F*-statistic for meta-regression model. All meta-regression models estimated using robust variance estimation to handle statistically dependent effect sizes. The analytic sample size was *n* = 47 studies and 111 effect sizes.

[a] See Farb and Margolis (2016).

[33] To make certain that study type (randomized controlled trial vs. quasi-experimental design) was not affecting the results, we ran all analyses on the sample of randomized controlled trials only. Results from these analyses, which are nearly identical to the main results reported here, are shown in Section 3.8 of the technical supplement.

7.2 Are Any Characteristics of Programs Associated with Better Attendance and Retention of Program Participants?

There were significant differences in participant attendance associated with group size, and significant differences in participant retention associated with some program components; but the other aspects of programs and characteristics of participants we tested were not significantly related to attendance and retention.

For this analysis we used performance data on attendance and retention for 51 and 50 studies, respectively. We fit a series of linear regression models to explore whether program design features, program implementation features, or participant characteristics were related to *participant attendance* (measured as the average percentage attendance rate across all program sessions) and *participant retention* (measured as the average proportion of participants attending 75 percent or more of the program sessions across all program periods).

Results are summarized in Exhibit 7-2 below, in a format similar to the summaries in previous sections.

How to Interpret Regression Coefficients in Exhibit 7-2

Because the outcome measures in Exhibit 7-2 are program percentages, the ***regression coefficients (in the column labeled "b") are in the metric of percentage points****. These are generally easier to interpret than log odds ratios.*

For example, in the first panel of results in Exhibit 7-2, the first column shows the association between program tier and attendance. This association was not statistically significant (p = .20). Nonetheless, the coefficients can be interpreted as follows: The intercept estimate indicates that the average attendance rate in Tier 2 programs was 83 percent, and the coefficient associated with Tier 1 programs indicates that average attendance was 6 percentage points lower in Tier 1 programs (i.e., 77 percent).

Several factors in isolation were associated with attendance and retention rates. However, because it is difficult to separate these elements from other factors, we caution against over-interpreting these findings. Attendance rates (column 1) had a significant association with group size, the presence of certain program components, and program length. In particular, programs with "Other (mixed individual/group)" group sizes had the lowest attendance rates (21 percentage points lower than in large groups), whereas programs delivered individually had the highest rates. Programs with condom demonstrations had significantly higher attendance rates than programs without condom demonstrations. Finally, programs with at least weekly contact had significantly lower attendance rates than programs without at least weekly contact. Retention rates (column 2) also had a significant positive association with condom demonstrations.

Attendance and retention rates were not significantly associated with any other observed attributes of program or participant characteristics, including *program focus*, *program length*, *gender specificity* (girls only), *group composition* (same gender or both genders), *program setting* (classroom, community, or other), *program delivery personnel* (health educators, classroom teachers, other), *implementation characteristics-fidelity*, or *participant characteristics*.

These attendance- and retention-related findings are difficult to interpret. It seems unlikely that the isolated characteristics identified here as being associated with attendance and retention (e.g., condom demonstrations) are truly the drivers of the programs' success in engaging or retaining participants. Rather, it is likely that these characteristics are correlated with other, unobserved factors that are responsible for these effects. As with the analysis of behavioral outcomes, it is also likely that better

attendance and retention are caused by complex combinations of factors rather than single factors in isolation. For these reasons, we suggest readers use caution when interpreting the isolated results in Exhibit 7-2.

Exhibit 7-2: Relationships between Program Design Features, Program Implementation Features, and Participant Characteristics to Attendance and Retention

	Attendance (k = 51)		Retention (k = 50)	
	b	95% Confidence Interval	*b*	95% Confidence Interval
Program Design Features				
Level of Prior Evidence (Program Tier)				
Tier 2 program	Ref.			
Tier 1 program	–6	[–15, 3]	–9	[–20, 3]
Intercept	83*	[77, 89]	79*	[72, 86]
	F = 1.70, p = .20		F = 2.38, p = .13	
Program Focus				
Sexual health	Ref.		Ref.	
Youth development	–8	[–17, 2]	–12	[–24, 1]
Other	12	[–3, 26]	12	[–6, 31]
Intercept	82*	[76, 87]	78*	[71, 85]
	F = 3.02, p = .06		F = 3.39, p = .04	
Program Components				
Condom demonstrations	13*	[2, 24]	15*	[1, 29]
Service learning	–10	[–29, 10]	–19	[–43, 5]
Role plays	–1	[–13, 14]	–2	[–15, 18]
Games	5	[–7, 16]	6	[–9, 20]
Reflective exercises	8	[–3, 19]	10	[–4, 23]
Direct provision of health services	–2	[–18, 14]	–3	[–24, 17]
Parent activities	12	[–0, 24]	14	[–1, 29]
Positive role model	3	[–16, 21]	6	[–17, 29]
Intercept	72*	[63, 81]	66*	[53, 78]
	F = 2.73, p = .02		F = 2.92, p = .01*	
Group Size				
Individualized	Ref.		Ref.	
Small groups (<10)	–1	[–18, 15]	–4	[–26, 17]
Large groups	–4	[–18, 9]	–6	[–24, 12]
Other (mixed individual/group)	–21*	[–42, –1]	–22	[–51, –6]
Intercept	85*	[73, 98]	82*	[66, 99]
	F = 1.88, p = .15		F = 0.88, p = .46	
Group Compositiona				
Youth separated into single-gender groups	2	[–8, 13]	6	[–7, 19]
Intercept	80*	[75, 85]	74*	[68, 81]
	F = 0.22, p = .64		F = 0.87, p = .35	

	Attendance (k = 51)		Retention (k = 50)	
	b	95% Confidence Interval	*b*	95% Confidence Interval
Gender Specificity (girls only)				
Program designed for girls only[b]	–3	[–14, 9]	1	[–15, 17]
Intercept	81*	[76, 86]	76*	[70, 82]
	F = 0.19, p = .66		F = 0.02, p = .89	
Program Length				
At least weekly contact	–12*	[–23, –1]	–13	[–27, 1]
Contact hours	–0	[–0, 0]	–0	[–0, 0]
Intercept	91*	[81, 101]	87*	[75, 100]
	F = 2.83, p = .07		F = 2.81, p = .07	
Program Implementation Features				
Program Setting				
Classroom	Ref.		Ref.	
Community	–7	[–18, 4]	–8	[–22, 6]
Other	–2	[–13, 9]	–2	[–16, 12]
Intercept	83*	[76, 89]	78*	[70, 86]
	F = 0.82, p = .45		F = 0.66, p = .52	
Program Delivery Personnel				
Health educators	–7	[–22, 8]	–10	[–29, 9]
Classroom teachers	Ref.		Ref.	
Other	–9	[–24, 6]	–12	[–31, 6]
Intercept	88*	[75, 101]	86*	[69, 102]
	F = 0.76, p = .47		F = 0.90, p = .42	
Implementation Characteristics–Fidelity				
Facilitator reports	64	[–14, 142]	98	[–0, 195]
Intercept	21	[–54, 95]	–16	[–1.09, 77]
	F = 2.74, p = .11		F = 4.05, p = .05	
Participant Characteristics (k = 35)				
Percentage boys	4	[–22, 31]	9	[–28, 45]
Percentage Black	–10	[–30, 9]	–12	[–40, 15]
Percentage Hispanic	5	[–16, 27]	6	[–24, 36]
Average age	4	[–0, 9]	5	[–1, 11]
Risk (control event rate)	–18	[–53, 18]	–18	[–67, 31]
Intercept	28	[–28, 84]	14	[–64, 91]
	F = 1.55, p = .20		F = 1.15, p = .36	

Notes. *b* = unstandardized meta-regression coefficients, *F* = omnibus *F*-statistic for meta-regression model, Ref. = reference category. All meta-regression models estimated using robust variance estimation to handle statistically dependent effect sizes. Regression coefficients converted to percentages by multiplying *100.

[a] Some programs or implementations deliver program content to mixed-gender groups, whereas others separate youth into single-gender groups.

[b] Gender-specific programs are designed to be delivered exclusively to youth of a single gender. Because program content is tailored for that gender, it would be inappropriate to deliver the program to youth of the other gender. All gender-specific programs were tailored for girls with the exception of Wise Guys, which was designed for boys only.

* p < .05.

8. Suggestive Findings

This chapter discusses a small number of results that could not be considered definitive because they did not meet our strict criteria for statistical significance, but that nonetheless suggest avenues for further exploration or development or provide additional context to the findings in Chapters 6 and 7. These results were close to, but did not exceed, traditional thresholds for statistical significance. As such**, these findings suggest areas in which future research may be warranted, but they should not be viewed as settling the open questions.**

Two program characteristics may have been associated with program effectiveness: gender specificity and delivery to individuals rather than to small groups.

Gender specificity. Of the 53 studies reviewed, 10 evaluated program models designed for all-girl samples: the *Safer Sex Intervention* (three studies), *17 Days*, *BUtiful*, *Will Power/Won't Power*, *AIM 4 Teen Moms*, *Go Grrrls*, *Steps to Success*, and *Teen Options to Prevent Pregnancy*.[34] We contrasted the effect of such programs with programs designed for adolescents of either gender. The results show that programs designed exclusively for girls were somewhat more effective ($b = 0.16$) than programs designed for both genders. However, the difference was not significant at the 95 percent confidence level ($p = .12$), and therefore we cannot with a high level of confidence reject that this difference might be due to chance. However, looking only at the 24 confirmatory effect sizes from the 10 studies with all-girl samples, we found an average effect size for *recent pregnancy* that was statistically significant ($p = .04$), indicating a small but favorable program effect on this behavior; we also found a marginally significant ($p = .06$) effect on participants' risky sexual behaviors overall.[35]

This finding pertains to gender-specific programs designed exclusively for girls. Such programs are distinct from mixed-gender programs that separate groups by gender when delivering the curriculum. We found no evidence that separating groups by gender in mixed-gender programs had an association with program effects ($p = .33$).

Individualized vs. small-group delivery. Programs were grouped into four categories based on their group size: individualized, small groups of less than 10, large groups of 10 or more, and mixed configurations. As described earlier, the model results show that these categories did not collectively explain a significant amount of variation in effect sizes ($p = .29$).

However, the contrast between individualized delivery and other types of delivery (small groups, large groups, and mixed) consistently favored individualized delivery, with *p*-values for these contrasts of approximately .13 to .15. The negative coefficients for small and large group and mixed delivery of $b = -0.25$ to -0.27 in Exhibit 6-1 indicate that, on average, programs that delivered services to groups had effect sizes that were somewhat smaller than did programs that delivered services to individuals.

It is perhaps not surprising that individualized delivery, single-gender targeting, and delivery in settings other than classrooms (which include clinics, participants' homes, online, and after-school) may have

---

[34] The *Steps to Success* evaluation sample was 100% girls; however, the program also offered services to their male partners.

[35] Looking at all effect sizes, not just for confirmatory outcomes, yielded similar results with *p*-values of 0.04 and 0.05 for recent pregnancy and overall, respectively, as well as a marginal significant effect on recent unprotected sexual activity (p=.07).

similarly favorable effects. These three characteristics have considerable overlap; that is, program models that deliver services in settings such as clinics, online, or in participants' homes tend to be implemented with individuals and are more likely to be targeted to girls than are, for example, classroom-based programs.

9. Discussion

This meta-analysis used advanced analytic methods to assess the effects of teen pregnancy prevention programs across a large collection of studies, focusing on the importance of various program design and implementation features. The group of federally funded evaluations that made up the sample was, in many ways, well suited for a meta-analysis. The individual evaluations were held to high quality design standards and they reported consistently measured outcomes and implementation characteristics. That is not to say, as we noted earlier, that there was not variation in the strength of program implementation and the quality of the actual evaluation, as it was implemented.

On the other hand, effect sizes for behavioral outcomes in most of the studies were modest and there was little variation in program effects across studies. Because a meta-regression aims to exploit variation in effects and search for explanatory factors, the combination of modest behavioral effects and little variation was an obstacle to that analysis.

9.1 Findings Summary

On average, these federally funded programs caused a small reduction in risky sexual behaviors and their consequences.

In particular, the programs had favorable effects on their most important (confirmatory) outcomes. These included a small, statistically significant overall effect across all outcomes and studies; it also included marginally significant effects on two specific behavioral outcomes: *recent pregnancy* and *ever had sex*. On average, the effect sizes were favorable but small. The effect on *ever had sex* translates into a decline from 19 percent in the control group (the average for studies that reported this outcome) to 18 percent in the treatment group. One exception is the outcome *recent pregnancy*, for which the overall effect size was somewhat larger—a decline from 16 percent in the control group (again, the average for studies that reported this outcome) to 12 percent in the treatment group.

Programs implemented in classrooms had less favorable effects than programs implemented in other settings such as clinics, after-school settings, participants' homes, or online.

We found a statistically significant relationship between program setting and program effects. We categorized programs according to whether the program setting was a classroom, community center, or some other setting (e.g., clinics, after-school settings, participants' homes, or online). In particular, programs implemented in classrooms produced significantly less favorable effects than did programs implemented elsewhere.

There are hints that programs may be more effective when designed exclusively for girls and when they deliver services to individuals rather than to groups of teens.

The findings that programs designed for girls and those delivered individually may be associated with greater program effects should be considered merely suggestive, because these results did not meet our criteria for statistical significance.

9.2 Interpretation

The finding that these HHS-funded programs, on average, caused a small reduction in risky sexual behaviors is promising news for the field.

The finding that classroom-based programs are less effective than programs implemented in other settings may also offer some direction for program developers and implementers. Hints that programs designed

for girls and those delivered individually may be linked with more favorable program effects may inform future research in this area. (Because of the rarity of programs that target boys specifically, we had no way to examine the effectiveness of a comparable set of programs aimed at them.) Based on these findings, it seems worthwhile to re-examine assumptions about the effectiveness of interventions intended for co-ed classrooms in changing behavioral outcomes.

Finally, though the number of new evaluations funded by HHS represents a dramatic expansion of the teen pregnancy prevention evidence base, it remains a relatively small sample for exploring between-study comparisons. As a result, most moderator analyses in this report were limited to the effect of a single variable. Given the complexity of the problem, including the social, psychological, and experiential antecedents, it is not surprising that there is not an obvious solution identifying a single factor in addressing teen pregnancy. It is unlikely that predicting program success is that simple. Rather, some combination of program and participant attributes (e.g., single-gender targeting, group size, setting, and age) may be required for a teen pregnancy prevention program to demonstrate effectiveness.

Fortunately, as additional well-designed studies become available for meta-analysis, our ability to investigate clusters of program and individual characteristics, and the interactions among them, will increase.

Appendix A. Study Identification and Selection

For the meta-analysis, we screened all evaluations of teen pregnancy prevention programs in the following categories:

1. Grantee-led evaluations conducted by TPP Program grantees through cooperative agreements with OAH;
2. Grantee-led evaluations conducted by PREIS Program grantees through cooperative agreements with FYSB;
3. OAH/ASPE-led TPP Replication Study;
4. OAH-led Adolescent Pregnancy Prevention Approaches (PPA) Study; and
5. ACF-led Personal Responsibility Education Program (PREP) Multi-Component Evaluation.

HHS provided all such reports completed (two of them in draft form) before April 15, 2019. We screened each report for eligibility using pre-specified criteria designed to ensure that findings were comparable across evaluations.

For Additional Technical Detail

The report's technical supplement provides substantial additional detail on study identification, selection, and coding. The supplement includes the full list of eligibility criteria (Section 1.1), data extraction and coding procedures for the aggregate data (Section 1.2), and instructions that were given to grantees on how to provide individual participant data (Section 1.3).

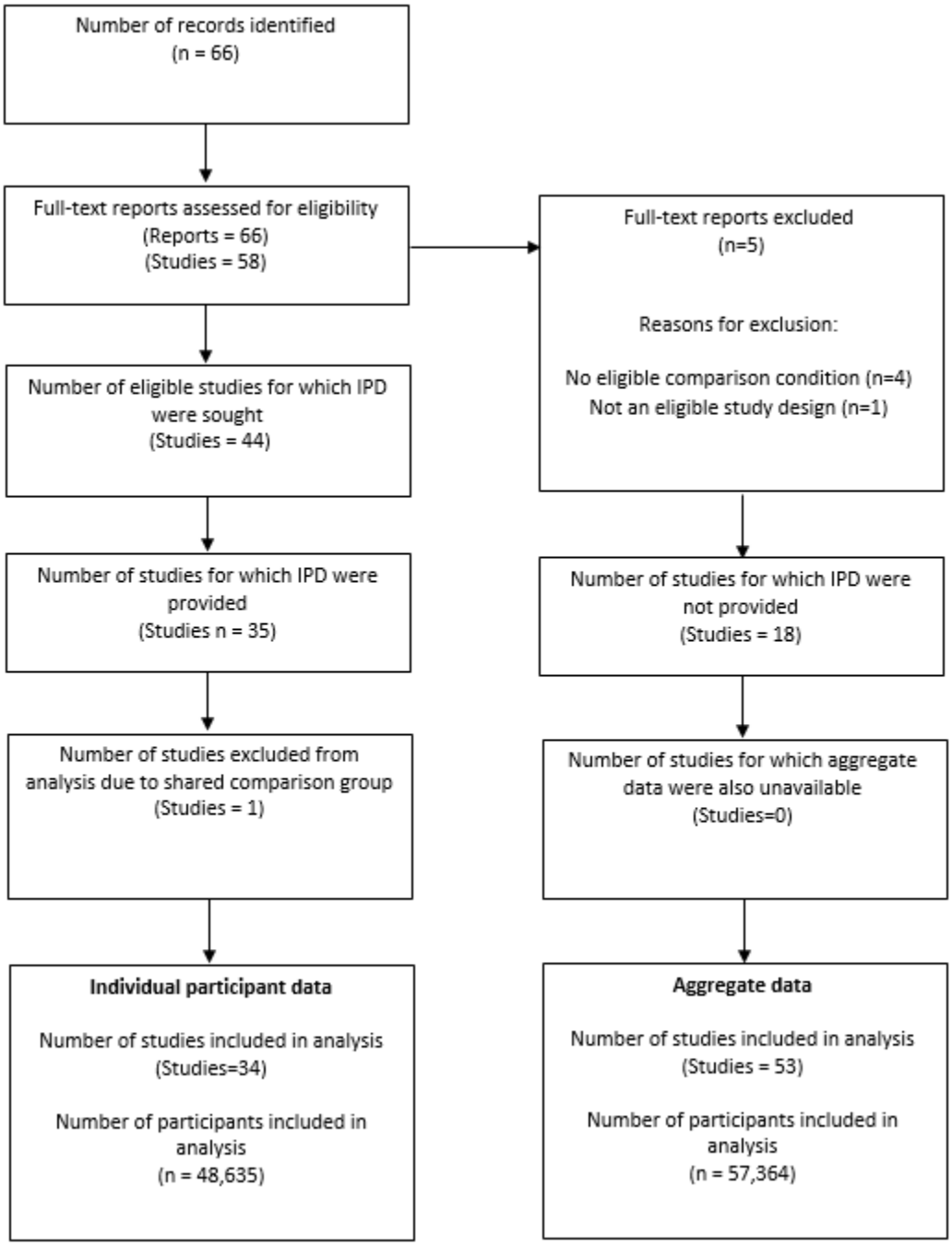
Studies were eligible for the meta-analysis if they used rigorous research methods and reported program impacts on youth sexual behavior, pregnancy, or sexually transmitted infections. In particular, the study must have assessed impacts using experimental or high-quality quasi-experimental research designs, comparing youth who received a teen pregnancy prevention program with youth who did not receive similar services. The study must have also had a sample of at least 10 study participants whose average age was between 10 and 19 years old. HHS allowed the Abt team to review all studies regardless of favorability or statistical significance, so that the presence or absence of significant findings was not a criterion for inclusion.

Exhibit A-1 describes the eligibility screening process.

We received 66 candidate reports from HHS. These included 33 final reports and one journal manuscript from OAH grantee-led evaluations, seven final reports from PREIS grantee-led evaluations; three interim reports and three final reports from the TPP Replication Study; six interim reports and five final reports from the PPA Study; and three interim reports and three final reports from the PREP Multi-Component Evaluation. Of these, 61 reports, describing the results of 53 independent studies, were determined to be eligible for the meta-analysis. These studies comprised 57,364 studied youth.

Individual participant data from the authors of many of the eligible evaluations were provided; in particular authors provided data for 35 (66 percent) of the eligible studies, 34 of which were included in the meta-analysis. These 34 study samples comprised 48,635 studied youth.

Exhibit A-1: Study Identification Flow Diagram



Appendix B. List of Studies Included in the Meta-Analysis

***Asterisks denote reports included in the aggregate data analysis but not in the individual participant data analysis.**

Abe, Y., L. Toms Barker, V. Chan, and J. Eucogco. 2016. *Early Findings from the Evaluation of the Pono Choices Program – A Culturally-Responsive Teen Pregnancy and Sexually Transmitted Infection Prevention Program for Middle School Youth in Hawai'i.* Columbia, MD: IMPAQ International.

Abt Associates. 2016a. *¡Cuídate!: Interim Impact Report.* Cambridge, MA: Abt Associates. Note: This report contributes three independent study samples (sites) to the meta-analysis.

Abt Associates. 2016b. *Reducing the Risk: Short-Term Impact Report.* Cambridge, MA: Abt Associates. Note: This report contributes three independent study samples (sites) to the meta-analysis.

Abt Associates. 2016c. *Safer Sex Intervention: Short-Term Impact Report.* Cambridge, MA: Abt Associates. Note: This report contributes three independent study samples (sites) to the meta-analysis.

Abt Associates. 2017a. *¡Cuídate!: Final impact report.* Unpublished manuscript. Note: This report contributes three independent study samples (sites) to the meta-analysis.

Abt Associates. 2017b. *Reducing the Risk: Longer-term impact report.* Unpublished manuscript, Abt Associates, Cambridge, MA. Note: This report contributes three independent study samples (sites) to the meta-analysis.

Abt Associates. 2017c. *Safer Sex Intervention: Short-term impact report.* Unpublished manuscript, Abt Associates, Cambridge, MA. Note: This report contributes three independent study samples (sites) to the meta-analysis.

Advanced Empirical Solutions. 2015. *Evaluation of Will Power/Won't Power in Los Angeles County.* Los Angeles, CA: Advanced Empirical Solutions, LLC.

Calise, T. V., W. Chow, and K. F. Dore. 2015. *Evaluation of Healthy Futures in Public Middle Schools in Three Northeastern Massachusetts Cities: Findings from an Innovative Teen Pregnancy Prevention Program.* Boston, MA: JSI Research and Training Institute, Inc.

Carter, S. L., B. Beadnell, and J. Vanslyke. 2015. *Evaluation of the Web of Life Teen Pregnancy Prevention Program: Findings from an Innovative Positive Youth Development Approach for American Indian Youth.* Albuquerque, NM: National Indian Youth Leadership Project.

*Covington, R. D., G. Goesling, C. Trenholm, J. Manlove, L. Welti, P. Drake, and J. Glassman. 2015. *Interim Impacts of the AIM 4 Teen Moms Program.* Princeton, NJ: Mathematica Policy Research.

*Covington, R. D., Goeslin, B., Tuttle, C. C., Crofton, M., Manlove, J., Oman, R. F., & Vesely, S. 2016. *Final impacts of the POWER through Choices program.* Washington, DC: Mathematica Policy Research

*Covington, R. D., Luca, D. L., Manlove, J., Welti, K. (2017). *Final Impacts of AIM 4 Teen Moms.* Washington, DC: U.S. Department of Health and Human Services, Office of Adolescent Health.

Covington, R., D., Wood, R., G., Goesling, B. (2019). Focusing on the Boys: The Longer-Term Impacts of Wise Guys in Davenport, Iowa.* Princeton, NJ: Mathematica Policy Research

Coyle, K. K., S. C. Potter, J. R. Glassman, L. McDade-Montez, and T. Unti. 2015. *Evaluation of It's Your Game ... Keep It Real in South Carolina: Final Report.* Scotts Valley, CA: ETR Associates.

Coyle, K., P. Anderson, B. A. Laris, T. Unti, H. Franks, and J. Glassman. 2016. *Evaluation of It's Your Game ... Keep It Real in Houston, TX: Final Report.* Scotts Valley, CA: ETR Associates.

Crean, H. F., S. M. Seibold-Simpson, M. Jambon, and R. E. Kreipe. 2016. *Evaluation of the Teen Outreach Program® in Rochester, New York: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program.* Rochester, NY: University of Rochester School of Nursing.

Cunningham, M. R., M. A. van Zyl, and K. W. Borders. 2016. *Evaluation of Love Notes and Reducing the Risk in Louisville, Kentucky.* Louisville, KY: University of Louisville Research Foundation. Note: This report contributes two study samples (sites) to the AD meta-analysis and contributes one study sample to the IPD meta-analysis.

Daley, E. M., E. R. Buhi, W. Wang, A. Singleton, R. Debate, S. Marhefka ... R. Ziemba. 2015. *Evaluation of Wyman's Teen Outreach Program® in Florida: Final Impact Report for Florida Department of Health. Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program.* Tallahassee, FL: Florida Department of Health.

Dierschke, N., J. Gelfond, D. Lowe, R. S. Schenken, and K. Plastino. 2015. *Evaluation of Need to Know (N2K) in South Texas: Findings from an Innovative Demonstration Program for Teen Pregnancy Prevention Program.* San Antonio, TX: The University of Texas Health Science Center at San Antonio.

Eichner, J., J. Salaway, J. Smith-Jones, and R. McCall. 2015. *Evaluation of Seventeen Days in Ohio, Pennsylvania, and West Virginia: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program.* Pittsburgh, PA: University of Pittsburgh, Office of Child Development.

Francis, K., M. Woodford, and M. Kelsey. 2015. *Evaluation of the Teen Outreach Program in Hennepin County, MN: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program.* Cambridge, MA: Abt Associates Inc.

*Fronius, T., Guckenburg, S., Petrosino, A., Taylor, D., Persson, H., MacDougall, P., Fuxman, S., & O'Donnell, L. (2016). *More Than a Dream/Mas Que Un Sueno: Final Evaluation Report.* Washington, DC: WestEd.

*Goesling, B., S. Colman, M. Scott, and E. Cook. 2014. *Impacts of an Enhanced Family Health and Sexuality Module of the HealthTeacher Middle School Curriculum*. Princeton, NJ: Mathematica Policy Research.

*Goesling, B., Covington, R. D., Manlove, J., Barry, M., Oman, R. F., & Vesely, S. 2015. *Interim impacts of the POWER Through Choices Program.* Princeton, NJ: Mathematica Policy Research.

*Goesling, B., Lee, J., Wood R., Zief, S. 2018. *Adapting an Evidence-based Curriculum in a Rural Setting: The Longer-Term Impacts of Reducing the Risk in Kentucky* (OPRE Report Number: 2018-27). Princeton, NJ: Mathematica Policy Research

*Goesling, B., Wood R., Covington, R. D. 2018. *Focusing on the Boys: The Early Impacts of Wise Guys in Davenport, Iowa* (OPRE Report Number: 2018-43). Princeton, NJ: Mathematica Policy Research

*Goesling, B., Wood, R.G., Lee, J., Zief, S. 2017. *Adapting an Evidence-based Curriculum in a Rural Setting: The Early Impacts of Reducing the Risk in Kentucky* (OPRE Report Number: 2017-43). Princeton, NJ: Mathematica Policy Research

Herrling, S. 2016. *Evaluation of the Children's Aid Society (CAS)-Carrera Adolescent Pregnancy Prevention Program in Chicago, IL: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program.* Accord, NY: Philliber Research and Evaluation.

Kissinger, P., N. Schmidt, J. Green, J. Latimer, A. Madkour, G. Clum, and C. Johnson. 2015. *Evaluation of BUtiful: An Internet Pregnancy Prevention for Older Teenage Girls in New Orleans, Louisiana.* New Orleans, LA: Tulane University School of Public Health and Tropical Medicine.

Kissinger, P., Schmidt, N., Green, J., Latimer, J., Madkour, A., Clum, G., & Johnson, C. 2017. *BUtiful: An internet pregnancy prevention intervention for older teenage women in New Orleans, Louisiana.* New Orleans, LA: Tulane University School of Public Health and Tropical Medicine.

*LaChausse, R. 2015. *Evaluation of the Positive Prevention PLUS Teen Pregnancy Prevention Program.* San Bernardino, CA: California State University.

*Lauby, J., Milnamow, M., Stup, E., Brown, J., Batson, H. 2017. *Final Evaluation Impact Report Plain Talk Philadelphia.* Philadelphia, PA: Public Health Management Corporation

*LeCroy & Milligan Associates, Inc. 2016. *Study of Go Grrrls in Pima County: Findings from an innovative teen pregnancy prevention program.* Tuscon, AZ: LeCroy & Milligan Associates, Inc.

*Martin, S., A. Hill, M. Nye, and K. Hollman-Billmeier. 2015. *Evaluation of Alaska Promoting Health Among Teens, Comprehensive Abstinence and Safer Sex (AKPHAT) in Alaska.* Anchorage, AK: Institute of Social and Economic Research, University of Alaska Anchorage.

Philliber, A. E., S. Philliber, and S. Brown. 2016. *Evaluation of the Teen Outreach Program® in the Pacific Northwest.* Accord, NY: Philliber Research and Evaluation.

Philliber, A. E., and S. Philliber. 2016. *Evaluation of the Teen Outreach Program® in Kansas City, Missouri.* Accord, NY: Philliber Research and Evaluation.

*Philliber, W., & Hirsch, H. 2016. *Evaluation of DFY: Take Charge!, An Adolescent Pregnancy Prevention Program of Cicatelli Associates Inc. (CAI): Findings from a Personal Responsibility Education Innovative Strategies Program.* Accord, NY: Philliber Research & Evaluation

Piotrowski, Z. H., and D. Hedeker. 2016. "Evaluation of the *Be The Exception* Sixth-Grade Program in Rural Communities to Delay the Onset of Sexual Behavior." *American Journal of Public Health* (Supplement): s132-s139.

Supplementary reports:

Piotrowski, Z. H., and D. Hedeker. 2015. E*valuation of Positive Potential Middle School Program in Rural Northwest Indiana Communities.* Chicago, IL: University of Chicago.

Piotrowski, Z. H., and D. Hedeker. 2015. *Evaluation of the Positive Potential Be the Exception Grade 6 Program in Predominantly Rural Communities.* Chicago, IL: University of Chicago.

Piotrowski, Z. H., & Hedeker, D. 2016. *Evaluation of Positive Potential Middle School Program in Rural Northwest Indiana Communities: Findings from an Innovative Teen Pregnancy Prevention Program.* Chicago, IL: University of Chicago.

Policy and Research Group, The. 2015. *Evaluation of Becoming a Responsible Teen: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program.* New Orleans, LA: Author.

Robinson, W. T., R. Kaufman, and L. Cahill. 2016. *Evaluation of the Teen Outreach Program in Louisiana*. New Orleans, LA: Louisiana State University Health Sciences Center at New Orleans, School of Public Health.

Rotz, D., B. Goesling, M. Crofton, C. Trenholm, J. Manlove, and K. Welti. 2016. *Interim Impacts of Teen PEP in New Jersey and North Carolina High Schools.* Cambridge, MA: Mathematica Policy Research.

Rotz, D., Goesling, B., Crofton, M., Manlove, J., & Welti, K. 2016. *Final impacts of Teen PEP (Teen Prevention Education Program) in New Jersey and North Carolina High Schools.* Washington, DC: Mathematica Policy Research.

Rotz, D., Lucia, D. L., Goesling, B., Cook, E., Murphy, K., Stevens, J. 2016. *Final Impacts of the Teen Options to Prevent Pregnancy Program.* Cambridge, MA: Mathematica Policy Research.

*Rotz, D., Shiferaw, M., & Wood, R. G. 2019. *Enhancing a Home Visiting Program to Address Repeat Adolescent Pregnancy: The Impacts of Steps to Success.* Princeton, NJ: Mathematica Policy Research.

*Rotz, D., Wood, R. 2018. *Enhancing a Home Visiting Program to Address Repeat Adolescent Pregnancy: The Early Impacts of Steps to Success.* Princeton, NJ: Mathematica Policy Research.

*Ruwe, M. B., L. McCloskey, A. Meyers, N. Prudent, and M. K. Foureau-Dorsinville. 2016. *Evaluation of Haitian-American Responsible Teen: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program in Eastern Massachusetts*. Boston, MA: Boston Medical Center.

*Schwinn, T., C. E. Kaufman, K. Black, E. M. Keane, N. R. Tuitt, C. K. Big Crow ... S. Schinke. 2015. *Evaluation of mCircle of Life in Tribes of the Northern Plains: Findings from an Innovative Teen Pregnancy Prevention Program.* Denver, CO: University of Colorado Denver, Centers for American Indian and Alaska Native Health.

*Seshadri, R., C. Smithgall, R. George, J. Ippolito, D. Dasgupta, E. Wiegand, S. Guiltinan, and M. Wood. 2015. *Evaluation of Teen Outreach Program in Chicago: Final Impact Report for Chicago Public Schools.* Chicago, IL: Chapin Hall at the University of Chicago.

Slater, H. M., and D. B. Mitschke. 2015. *Evaluation of the Crossroads Program in Arlington, TX: Findings from an Innovative Teen Pregnancy Prevention Program.* Arlington, TX: University of Texas at Arlington.

Smith, K. V., C. Dye, E. Cook, K. Rosinksy, and M. Scott. 2016. *Interim Impacts of the Gender Matters Program.* Princeton, NJ: Mathematica Policy Research.

Smith, K. V., Dye, C., Rotz, D., Cook, E., Rosinsky, K., & Scott, M. 2016. *Final impacts of the Gender Matters Program.* Washington, DC: Mathematica Policy Research.

Smith, K., D. Rotz, B. Goesling, E. Cook, K. Murphy, and J. Stevens. 2015. *Interim Impacts of the Teen Options to Prevent Pregnancy Program.* Princeton, NJ: Mathematica Policy Research.

*Tanner, J. F., Komarek, T., Manolis, C., Fairall, G., Smith, L., & Tucker, E. 2016. *Evaluation of Choosing the Best and Possessing Your Power: Comparison of an evidence-based teen pregnancy prevention program and to a new program.* Norfolk, VA: JK Tanner Inc.

*Tucker, T. 2015. *Evaluation of the Carrera Program: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program.* Atlanta, GA: Tressa Tucker and Associates, LLC.

*Usera, J. J., and K. M. Curtis. 2015. *Evaluation of the Ateyapi Identity Mentoring Program in South Dakota: Findings from the Replication of an Evidenced-Based Teen Pregnancy Prevention Program.* Sturgis, SD: Delta Evaluation Consulting, LLC.

Vyas, A., S. Wood, M. Landry, G. Douglass, and S. Fallon. 2015. *The Evaluation of the Be Yourself / Sé Tú Mismo in Montgomery and Prince George's Counties, Maryland.* Washington, DC: The George Washington University Milken Institute School of Public Health.

Walker, E. M., R. Inoa, and N. Coppola. 2016. *Evaluation of Promoting Health Among Teens! Abstinence-Only Intervention in Yonkers, NY.* Princeton, NJ: Sametric Research.

Appendix C. Excluded Studies, with Reasons for Exclusions

Of the 66 reports we received from the Office of Adolescent Health, we deemed five to be ineligible:

1. Bull, S., S. Schmiege, and S. Devine. 2015. *Evaluation of Youth All Engaged (YAE) in Denver, CO*. Denver, CO: Denver Public Health Department.

   Reason: Compares two active teen pregnancy prevention programs; no eligible comparison condition.

2. Drake, A., Coman, E., Parikh, B., Mogro-Wilson, C., Martin-Peele, M., & Fifield, J. 2016. *Evaluation of the FatherWorks program*. Farmington, CT: UConn Health Disparities Institute.

   Reason: Compares two active pregnancy prevention programs; no eligible comparison condition.

3. Jenner, E., S. Walsh, L. W. Jenner, H. Demby, and A. Gregory. 2015. *Evaluation of Safer Sex Intervention in New Orleans, LA: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program*. New Orleans, LA: The Policy & Research Group.

   Reason: Reports only findings on sexual knowledge and attitude outcomes; no eligible sexual behavior/STI outcome.

4. LeCroy & Milligan Associates, Inc. 2016. *Teen Outreach Pregnancy Services PSPS Project final report*. Tucson, AZ: LeCroy & Milligan Associates, Inc.

   Reason: Compares two active conditions where comparison services are not business as usual.

5. Usera, J. J., and K. M. Curtis. 2015. *Evaluation of the Aban Aya Youth Development Project in the Mississippi Delta: Findings from the Replication of an Evidence-Based Teen Pregnancy Prevention Program*. Sturgis, ND: Delta Evaluation Consulting, LLC.

   Reason: Quasi-experimental design with matching only on age, race, and gender; no eligible randomized controlled trial or quasi-experimental design with controls for baseline sexual behavior.

Appendix D. Intervention, Comparison Condition, and Confirmatory Outcome(s) for Eligible Studies, by Report

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Abe et al. (2016)	*Pono Choices*	HI	Participants received a total of 9.5 hours of instruction over 10 sessions on sexual risk reduction. Program content covered goal setting, relationships, communication, knowledge about STIs, contraception, correct condom use, and refusal skills.	Business as usual. Participants attended existing health classes and may have covered topics such as reproductive anatomy, pregnancy and STI prevention, refusal skills, and condom use.	Engagement in high-risk sexual behavior
Abt Associates (2016a)	*¡Cuídate!*	AZ	Participants received eight 60-minute modules on reducing risky sexual behavior. Program content covered preventing HIV/AIDS and pregnancy, condom use, and refusal skills, while incorporating Latino values.	Healthy lifestyle curriculum. Participants received eight 60-minute sessions on self-esteem, body image, nutrition, exercise, stress, and the consequences of substance use.	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days
Abt Associates (2017a)	*¡Cuídate!*	AZ	(Longer-term follow up report for Abt Associates 2016a)	(Longer-term follow up report for Abt Associates 2016a)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016a)	*¡Cuídate!*	CA	Participants received eight 60-minute modules on reducing risky sexual behavior. Program content covered preventing HIV/STIs, condom use, and refusal skills, while incorporating Latino values.	Business as usual. Participants attended existing physical education classes.	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days
Abt Associates (2017a)	*¡Cuídate!*	CA	(Longer-term follow up report for Abt Associates 2016a)	(Longer-term follow up report for Abt Associates 2016a)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016a)	*¡Cuídate!*	MA	Participants received 6 total hours of instruction on reducing risky sexual behavior. Program content covered preventing HIV/AIDS, condom use, and refusal skills, while incorporating Latino values.	Business as usual. Existing programs varied by site but mainly consisted of regular physical education or health classes, after-school activities, or youth sport clubs.	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Abt Associates (2017a)	*¡Cuídate!*	MA	(Longer-term follow up report for Abt Associates 2016a)	(Longer-term follow up report for Abt Associates 2016a)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016b)	*Reducing the Risk (RtR)*	CA	Participants received sixteen 45-minute sessions on sexual health. Program content covered STI/HIV prevention, birth control methods, abstinence, healthy relationships, and refusal skills.	Business as usual. Participants attended usual physical education, health, or science classes, depending on the school.	Currently sexually active (in last 90 days) Sexual intercourse without birth control (in last 90 days)
Abt Associates (2017b)	*Reducing the Risk (RtR)*	CA	(Longer-term follow up report for Abt Associates 2016b)	(Longer-term follow up report for Abt Associates 2016b)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016b)	*Reducing the Risk (RtR)*	IL MO	Participants received sixteen 45-minute sessions on sexual health. Program content covered STI/HIV prevention, birth control methods, abstinence, healthy relationships, and refusal skills.	Business as usual. Participants attended the usual physical education, ROTC, health, or homeroom/guidance classes, depending on the school.	Currently sexually active (in last 90 days) Sexual intercourse without birth control (in last 90 days)
Abt Associates (2017b)	*Reducing the Risk (RtR)*	IL MO	(Longer-term follow up report for Abt Associates 2016b)	(Longer-term follow up report for Abt Associates 2016b)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016b)	*Reducing the Risk (RtR)*	TX	Participants received sixteen 45-minute sessions on sexual health. Program content covered anatomy, STI/HIV prevention, birth control methods, abstinence, healthy relationships, and refusal skills.	Business as usual. Participants attended nine 90-minute sessions of health class.	Currently sexually active (in last 90 days) Sexual intercourse without birth control (in last 90 days)

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Abt Associates (2017b)	*Reducing the Risk (RtR)*	TX	(Longer-term follow up report for Abt Associates 2016b)	(Longer-term follow up report for Abt Associates 2016b)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016c)	*Safer Sex Intervention (SSI)*	FL	Participants received one 55-minute and three 15-minute sessions on practicing safe sex. Program content covered preventing pregnancy and STIs, birth control, consequences of risky sex, and talking about sex with your partner.	Business as usual. Participants received the existing clinic standard-of-care, as well as pregnancy and STI tests.	Currently sexually active (in last 90 days) Sexual intercourse without birth control (in last 90 days)
Abt Associates (2017c)	*Safer Sex Intervention (SSI)*	FL	(Longer-term follow up report for Abt Associates 2016c)	(Longer-term follow up report for Abt Associates 2016c)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016c)	*Safer Sex Intervention (SSI)*	MN	Participants received one 55-minute and three 15-minute sessions on practicing safe sex. Program content covered preventing pregnancy and STIs, birth control, consequences of risky sex, and talking about sex with your partner.	Business as usual. Participants received the existing clinic standard-of-care.	Currently sexually active (in last 90 days) Sexual intercourse without birth control (in last 90 days)
Abt Associates (2017c)	*Safer Sex Intervention (SSI)*	MN	(Longer-term follow up report for Abt Associates 2016c)	(Longer-term follow up report for Abt Associates 2016c)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Abt Associates (2016c)	*Safer Sex Intervention (SSI)*	TN	Participants received one 55-minute and three 15-minute sessions on practicing safe sex. Program content covered preventing pregnancy and STIs, birth control, consequences of risky sex, and talking about sex with your partner.	Business as usual. Participants received the existing clinic standard-of-care.	Currently sexually active (in the last 90 days) Sexual intercourse without birth control (in last 90 days)

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Abt Associates (2017c)	*Safer Sex Intervention (SSI)*	TN	(Longer-term follow up report for Abt Associates 2016c)	(Longer-term follow up report for Abt Associates 2016c)	Currently sexually active in the last 90 days Sexual intercourse without birth control in the last 90 days Pregnant or gotten someone pregnancy since baseline
Advanced Empirical Solutions (2015)	*Will Power / Won't Power (WPWP)*	CA	Participants received eleven 90-minute sessions of an abstinence-plus curriculum. Program content covered anatomy, hygiene, contraceptives, skill development, and decision-making.	Equal Earners, Savvy Spenders (EESS). Participants received ten 90-minute after- school sessions on economic literacy and managing personal finances. Topics included loans, investments, credit cards, and global economics.	Pregnancy incidence Sexual activity onset
Calise et al. (2015)	*Healthy Futures*	MA	Participants received eight 50-minute classroom modules on relationship education each year for a total of 24 sessions. Supplementary services included virtual classrooms, parent websites, parent workshops, an after-school youth leadership program, and a summer youth program.	General health education curriculum. Participants received two 50-minute classroom modules on general health education topics each year (e.g., puberty, reproduction, bullying prevention, dating-violence prevention, and mental health promotion).	Ever had vaginal sex
Carter et al. (2015)	*The Web of Life (WOL)*	NM	Participants received 26 classroom, 20 after-school, 10 day-long off-site, and 3 multi-day overnight sessions throughout the school year, totaling 261 hours. Program content was delivered experientially through adventure-based programming and service learning, and included hiking, climbing, and caving, as well as education about healthy relationships, communication skills, and interpersonal awareness.	Business as usual. Participants received standard school curriculum, which typically included some sessions on sexual and reproductive health.	Having sexual intercourse
Covington et al. (2017)	*AIM 4 Teen Moms*	CA	Participants received seven 60-minute and two 90-minute sessions on the promotion of healthy lifestyles and behaviors. Program content covered planning for the future, reproductive planning, and parenting ideals.	Business as usual. Participants had access to existing community programs and services designed for teen mothers.	Repeat pregnancy

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Covington et al. (2016)	*POWER Through Choices (PTC)*	CA MD OK	Participants received ten 90-minute sessions on teen pregnancy, HIV, and STI prevention. Program emphasized self-empowerment and the impact of choices.	Business as usual. Participants had access to existing community and group home services but received limited or no sexual and reproductive health education or services.	Had sex without using any effective method of protection
Covington et al. (2019)	*Wise Guys*	IA	(Longer-term follow up to Goesling et al. 2018)	(Longer-term follow up to Goesling et al. 2018)	Ever had sexual intercourse
Coyle et al. (2015)	*It's Your Game … Keep It Real (IYG)*	SC	Participants received twenty-four 50-minute sessions on life skills and decision-making. Program content covered creating personal limits for risky behavior and developing avoidance strategies.	Business as usual. Programs varied by school district. Content included state-mandated information on reproductive health and STIs.	Initiation of sex (among virgins at baseline)
Coyle et al. (2016)	*It's Your Game … Keep It Real (IYG)*	TX	Participants received twenty-four 50-minute sessions on life skills and decision-making. Program content covered creating personal limits for risky behavior and developing avoidance strategies.	Business as usual. Participants received the standard health and sex education program. Programs varied by school district.	Initiation of vaginal or oral sex
Crean et al. (2016)	*Teen Outreach Program (TOP)*	NY	Participants received a minimum of twenty-five 45-minute sessions on positive youth development, with an additional 20 hours of community service learning. Program content covered goal setting, decision-making, autonomy, and healthy interactions with adults and peers.	*Work Readiness (WR)* curriculum. Participants received 60- to 90-minute monthly sessions for 9 months on topics such as customer service skills, career preparation, communication, and developing leadership skills.	Ever had sexual intercourse
Cunningham et al. (2016)	*Love Notes (LN)*	KY	Participants received 2 days of programming totaling 15 hours, on healthy relationships. Program content covered safety, communication strategies, problem solving, decision-making, sexuality, and domestic violence.	*The Power of We (POW)*. Participants received two weekends of programming totaling 15 hours on community-building and bringing about positive change in their local neighborhoods.	Number of partners past 3 months Sex without birth control past 3 months Sex without condom past 3 months
Cunningham et al. (2016)	*Reducing the Risk (RtR)*	KY	Participants received 2 weekends of programming totaling 15 hours on abstinence and safe sex. Program content covered the value of abstinence, refusal skills, delay tactics, preventing HIV/STIs, and using protection.	*The Power of We (POW)*. Participants received two weekends of programming totaling 15 hours on community-building and bringing about positive change in their local neighborhoods.	Number of partners past 3 months Sex without birth control past 3 months Sex without condom past 3 months

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Daley et al. (2015)	*Teen Outreach Program (TOP)*	FL	Participants received twenty-five 45-minute sessions on positive youth development, with an additional 20 hours of community service learning. Program content covered goal setting, communication and assertiveness, sexuality, and human development.	Business as usual health/physical education class. Participants were enrolled in usual school physical education or physical fitness class. Content covered healthy lifestyle choices, health and fitness, problem solving, and leadership skills.	Ever been pregnant or ever gotten someone pregnant Ever had sexual intercourse
Dierschke et al. (2015)	*Need to Know (N2K)*	TX	Participants received forty-eight 25-minute lessons on health and sexual education with additional reinforcement through social media and web content. Program content covered goal setting, risk behaviors, communication, STIs, abstinence, contraception, and dating violence.	Business as usual. There was no mandated sex education or pregnancy prevention program. Participants received standard health education during the school day.	Engaged in risky sexual behavior Ever had sex
Eichner et al. (2015)	*Seventeen Days*	OH PA WV	Participants watched an interactive video consisting of 2.5 hours of content, administered at a health clinic and continuing at their homes. Program content covered sexual negotiation, correct condom use, female anatomy, and STI information.	*Driving Skills for Life.* Participants watched an interactive video consisting of 2.5 hours of content, administered at a health clinic and continuing at their homes. Program content covered safe driving behaviors, such as how to handle a car in various road conditions.	Abstinent in past 3 months Unsafe sex behavior in the past 3 months
Francis et al. (2015)	*Teen Outreach Program (TOP)*	MN	Participants received a minimum of twenty-five 45-minute sessions on positive youth development, with an additional 20 hours of community service learning. Program content covered skill development, sexual health, and sexual behavior choices.	Business as usual. Participants received standard classroom curriculum (e.g., social studies, study hall, health/physical education), which varied across sites. Participants may also have received sex education as part of a health class or had guest speakers who addressed the topic of sexual health.	Recent sexual activity
Fronius et al. (2016)	*¡Cuídate!*	CO FL MO NM	Participants received six 60-minute modules of instruction on reducing risky sexual behavior among Latino/a youth. Program content covered abstinence, condom use, prevention of teen pregnancy, HIV and other STIs.	Active comparison condition. Participants attended a supervised activity for several hours. Time in the supervised activities varied by site.	Any sexual intercourse

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Fronius et al. (2016)	*More than a Dream / Más Que un Sueño*	CO FL MO NM	Participants received six 60-minute modules of *¡Cuídate!* and *Salud y Éxito* programs. *¡Cuídate!* was a culturally based program focused on sexual risky behaviors among Latino/a youth and covered abstinence, condom use, prevention of teen pregnancy, HIV and other STIs. *Salud y Éxito* was an audio-based health intervention for Latino/a parents that emphasized increased parent-child communication, rule setting, and monitoring related to risky sexual behaviors.	Active comparison condition. Participants attended a supervised activity for several hours. Time in the supervised activities varied by site.	Any sexual intercourse
Fronius et al. (2016)	*Salud y Éxito*	CO FL MO NM	Salud y Éxito was an audio-based health intervention delivered to Latino/a parents focused on delayed sexual initiation and pregnancy. Participants received a series of 3–5 minute community-informed, dramatic, gender- and developmentally crafted audio stories. Content covered increased parent-child communication, rule setting, and monitoring related to risky sexual behaviors.	Active comparison condition. Participants received monthly bilingual cards and booklets about nutrition, diet, and exercise geared toward parents of Latino/a children and adolescents.	Any sexual intercourse
Goesling et al. (2014)	*HealthTeacher (modified)*	IL	Participants received twelve 45- to 90-minute sessions on family health and sexuality. Program content covered anatomy, STD/HIV prevention, refusal skills, contraceptive methods, and sexual orientation.	Business as usual. Participants did not receive any sex education during the school year, though students were free to access community programs covering similar topics.	Ever had oral sex Ever had sexual intercourse
Goesling et al. (2018)	*Reducing the Risk (RtR) (adapted version)*	KY	Participants received an adapted 8-hour version of the teen pregnancy prevention curriculum. Program covered abstinence, contraception, STIs, and developing skills to avoid risky situations	Standard health curriculum. Participants received four class periods of sex education that did not include instruction on skills for avoiding sexual risk behaviors.	Had sexual intercourse in the past 3 months Had sexual intercourse without a condom in the past 3 months
Goesling et al. (2018)	*Wise Guys*	IA	Participants received fourteen 40–60 minute sessions on teen pregnancy prevention. Program content covered human sexuality, pregnancy, and the transmission of STIs.	Active business as usual. Participants received sexuality and reproductive health education provided as part of the regular school curriculum.	Ever had sexual intercourse

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Herrling (2016)	*CAS-Carrera*	IL	As part of a daily after-school program, participants received one 50-minute session per week focused on pregnancy prevention for 39 weeks. The prevention component focused on helping youth develop goals, broaden their sexual literacy, and increase awareness of the consequences of sexual activity. Additional services included academic and employment support, mental health and medical care, and sports and arts activities.	The Children's Home + Aid Community Schools Program. The after-school program offered a minimum of 29 hours of programming per month providing academic support, enrichment activities, and socio-emotional services.	Ever had sexual intercourse Use of contraception in the past 3 months
Kissinger et al. (2015)	*Be yoU, Talented, Informed, Fearless, Uncompromised and Loved (BUtiful)*	LA	Participants received eight 30-minute internet-delivered sessions on sexual health and pregnancy prevention. Program content covered contraception, pregnancy, STIs, and relationships.	*Diversity, Individuality, Vitality, Activity, and Strong (DIVAS).* Participants received eight 30-minute internet-delivered sessions on nutrition and wellness.	Use of reliable contraceptive method
LaChausse (2015)	*Positive Prevention PLUS*	CA	Participants received eleven 45-minute sessions on sexual health and pregnancy prevention. Program content covered healthy relationships, condom and contraception use, HIV/STI prevention, and refusal skills.	Business as usual. Participants received standard health education, science, or physical education. Instructors were not allowed to discuss pregnancy or STD prevention.	Ever been pregnant Ever had sex Had sex without birth control in prior 3 months
Lauby et al. (2017)	*Plain Talk Philadelphia*	PA	Participants received a community-level, multi-component intervention focused on sexual and reproductive health. Program included 11 two-hour sessions delivered to youth, 4-hour session for caregivers, and social media campaign. Content covered safe sex strategies, sexual risk behaviors, life skills and communication skills.	Business as usual. Adolescents participated in sexual health education routinely offered by schools, health care providers, and institutions in their communities.	Number of partners in the last three months Unprotected vaginal sex episodes in the past 3 months
LeCroy & Milligan Associates, Inc. (2016).	*Go Grrrls*	AZ	Participants received a 16-hour pregnancy prevention curriculum. Program covered gender role identity, positive body image, and self-image, responsible decision-making, healthy sexuality, making and keeping friends, accessing resources when needed, and planning for the future.	*TECHGyrls.* Participants received a 16-hour curriculum that focused on building academic competence and educational success.	Engaging in sexual intercourse

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Martin et al. (2015)	*Alaska Promoting Health Among Teens (AKPHAT)*	AK	Participants received twelve 60-minute sessions on sexual health and decision-making. Program content covered negotiation skills, contraceptive use and abstinence, and information on HIV, STIs, and pregnancy.	Business as usual. Participants had access to the existing school and community services.	Sex without condom in the past 3 months Sexual intercourse in the past 3 months
Philliber and Hirsch (2016)	*Development for Youth (DFY): Take Charge!*	NY	Participants received 12 two-hour sessions on pregnancy prevention and contraception use. Program covered self-efficacy, communication, feelings, decision-making, goal setting, sexual expression, physiology, reproduction, contraception, relationships, and sexually transmitted infections.	Active comparison condition (Healthy Living). Participants received a 2-hour program focused on the development of individual action plans for reducing health risks.	Recent unprotected sex
Philliber and Philliber (2016)	*Teen Outreach Program (TOP)*	MO	Participants received a minimum of twenty-five 45-minute sessions on positive youth development, with an additional 20 hours of community service learning. Program content covered values clarification, healthy relationships, communication, goal setting, decision-making, critical thinking, and human development/sexuality.	Business as usual. Participants received the existing curriculum in core classes (e.g., social studies, history). Schools typically offered health education, which did not cover reproductive health.	Ever had sexual intercourse Lack of recent birth control use
Philliber, Philliber, and Brown (2016)	*Teen Outreach Program (TOP)*	AK ID MT OR WA	Participants received a minimum of twenty-five 45-minute sessions on positive youth development, with an additional 20 hours of community service learning. Program content covered values clarification, relationships, communication and assertiveness, influence, goal setting, decision-making, and human development/sexuality.	*Community Voices (CV)* program. Participants attended four 60-minute sessions where they discussed current issues among young people in their communities. The program did not address sexuality or include service learning opportunities.	Ever been pregnant or caused someone to be pregnant
Piotrowski and Hedeker (2016)	*Positive Potential Be the Exception*	IN	Participants received five 45- to 50-minute sessions on risk avoidance and health promotion strategies. Program content covered maintenance of knowledge, skills, and behaviors to plan for a healthy future.	Business as usual. Participants attended standard health education classes, after-school activities, and community instruction about risk behaviors and health. Participants may also have attended assemblies with guest speakers who addressed general health and exercise.	Ever had sexual intercourse

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Robinson et al. (2016)	*Teen Outreach Program (TOP)*	LA	Participants received a minimum of twenty-five 45-minute sessions on positive youth development, with an additional 20 hours of community service learning. Program content covered values clarification, relationships, communication and assertiveness, influence, goal setting, decision- making, and human development/sexuality.	Business as usual. Participants received no programming but may have been exposed to similar programs in their communities.	Sex no birth control past 3 months
Rotz et al. (2016)	*Teen Prevention Education Program (Teen PEP)*	NC NJ	Participants received six 90-minute sessions on sex education. Program content covered postponing sexual activity, preventing pregnancy and HIV/STIs, and the consequences of using substances while engaging in sexual activity.	Business as usual. Participants received existing sexual health education programs, which varied across schools and across states.	Had sexual intercourse in the 3 months before survey Had sexual intercourse without a condom in the 3 months before the survey
Rotz et al. (2016)	*Teen Options to Prevent Pregnancy (T.O.P.P.)*	OH	(Long-term follow up to Smith et al. 2015)	(Long-term follow up to Smith et al. 2015)	Repeat pregnancy in past 18 months
Rotz et al. (2019)	*Steps to Success*	TX	Families received home visits weekly for 3 to 6 months and biweekly and monthly for up to 2 years. Program content covered parenting, child development, healthy birth spacing, father involvement, and mother's education and career planning.	Business as usual. Participants received home visits that focused on parenting and child development.	Repeat pregnancy
Ruwe et al. (2016)	*Haitian-American Responsible Teen (HART)*	MA	Participants received a 10-lesson educational intervention on sexuality and behavioral skills, as well as awareness of posttraumatic stress disorder.	Nutrition/fitness and PTSD awareness curriculum. Participants received a 10-lesson curriculum on promoting healthy eating habits, increasing physical activity, and awareness of posttraumatic stress disorder.	Ever had sex Had sex in the last 3 months Recent sex without condom Recent sex without effective birth control use
Schwinn et al. (2015)	*Multimedia Circle of Life (mCOL)*	ND SD	Participants received seven 60-minute class sessions and seven 25-minute online sessions on sexual risk reduction. Program content covered goal setting, decision-making, peer pressure, and information on HIV and STIs.	*After-School Science Plus (AS+).* Participants received seven 60-minute after-school sessions on physical science topics, such as bubbles, states of matter, and gravity.	Ever had sex

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Seshadri et al. (2015)	*Teen Outreach Program (TOP)*	IL	Participants received a minimum of twenty-five 45-minute sessions on positive youth development, with an additional 20 hours of community service learning. Program content covered skill development, sexual health, and sexual behavior choices.	General sexual health classes (i.e., business as usual). Participants received a minimum of 600 minutes of comprehensive sexual health education. Topics included reproduction, contraception, abstinence, and healthy sexual decision-making.	Had sex without a condom in the last 3 months
Slater and Mitschke (2015)	*Crossroads*	TX	Participants received 18.75 hours of programming over 3 consecutive days on sexual health. Program content covered building relationships, pregnancy and STI prevention, and identifying community resources.	Business as usual. Participants did not receive any comprehensive alternative sexual health program. They received drop-out prevention services as usual.	Vaginal intercourse without a condom
Smith et al. (2015)	*Teen Options to Prevent Pregnancy (TOPP)*	OH	Participants received eighteen 25-minute sessions on sexual health. Program content covered the use of effective contraception, knowledge and attitudes toward pregnancy prevention, birth spacing, and planning for the future. The program also provided access to a social worker and transportation to obtain contraceptive services.	Business as usual. Participants had access to existing reproductive health services in the local area.	No confirmatory outcome identified
Smith et al. (2016)	*Gender Matters (GEN.M)*	TX	Participants received five 4-hour sessions on gender and reproductive health. Program content covered understanding gender, healthy relationships, and pregnancy/STI prevention.	Business as usual. Participants were not offered any alternative programming, but had access to existing reproductive health services in the area.	Ever had sex Had sex without a condom in past 3 months Had sexual intercourse in past 3 months Had unprotected sex in past 3 months
Tanner et al. (2016)	*Choosing the Best (CTB)*	VA	Participants received six 2-hour educational sessions focused on abstinence. Program content covered information on STDs, HIV-AIDS, teen pregnancy, relationship education, refusal skills, and character education. Student manual provided at-home opportunities for parent-student interaction.	*Discovery.* Participants received six 2-hour sessions of a college preparation, career planning, and youth entrepreneurship program.	Risky sexual behavior

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Tanner et al. (2016)	*Possessing Your Power (PYP)*	VA	Participants received six 2-hour sessions on youth character development, enhancement, and empowerment. Program targeted seven risky behaviors including drugs, alcohol use, sex before marriage, tobacco, violence and crime, pornography, and gambling.	*Discovery*. Participants received six 2-hour sessions of a college preparation, career planning, and youth entrepreneurship program.	Risky sexual behavior
Policy & Research Group (2015)	*Becoming a Responsible Teen (BART)*	LA	Participants received eight 105-minute sessions on risky sexual behavior. Program content covered skill building (e.g., correct condom use, communication, and refusal techniques), values clarification, and intentions and attitudes toward sex.	*Healthy Living.* Participants received eight 1.75-hour sessions on healthy lifestyle choices, with a module on HIV from the BART intervention. The remaining seven modules covered topics such as nutrition, healthy eating, body image, and exercise.	Inconsistency of condom use
Tucker (2015)	*CAS-Carrera*	GA	As part of a daily after-school program, participants received one session per week on comprehensive health and sex education. The sex education component focused on anatomy, sexuality and reproduction, abstinence and contraception, STIs, and skills relevant to sexual behavior. Additional services included academic and employment support, mental health and medical care, and sports and arts activities.	The Boys and Girls Club (BGC). Participants received varied content from Club core components, such as career development, healthy lifestyles, leadership development, arts and culture, and sports and recreation. The *Smart Moves* curriculum was used to teach healthy lifestyles, including 12 sessions on preventing substance abuse and three sessions on early sexuality.	Ever had sex Sex without a condom or other birth control
Usera and Curtis (2015)	*Ateyapi Identity Mentoring Program*	SD	Participants received sixteen 50-minute Lakota-culture-based sessions on risk reduction behaviors and mentoring during and after school. Program content covered career exploration, goal setting, communication and decision-making, human sexuality, birth control methods, and STI/HIV prevention.	Mentoring. Participants did not receive any classroom-based lessons from the Ateyapi program, but had access to the same group of mentors used in the intervention.	Had recent sexual intercourse Safe sex

Study	Program Name	State(s)	Intervention	Comparison Condition	Confirmatory Outcome(s)
Vyas et al. (2015)	*Be Yourself / Sé Tú Mismo*	MD	Participants received at least nineteen 90-minute sessions of an after-school youth development curriculum on identifying and strengthening assets among Latino youth. Participants also took part in a weekend retreat, and received social media outreach and individual case management.	*Healthy Living / Vida Sana*. Participants received twelve 90-minute sessions on health- related topics such as fitness, nutrition, and exercise. The program included case management services and a 1-day weekend activity.	Contraceptive use at last sex (% yes) No contraceptive use in last 3 months (% yes) Sexual debut (% yes)
Walker et al. (2016)	*Promoting Health Among Teens! Abstinence-Only (PHAT-AO)*	NY	Participants received eight 60-minute modules on abstinence delivered on two consecutive Saturdays. Program content covered puberty, abstinence, HIV/STIs, pregnancy, and refusal skills.	*Promoting Health Among Teens! Health Intervention.* Participants received eight 60-minute modules on healthy behaviors delivered on two consecutive Saturdays. Content covered exercise, healthy eating, stress, and substance use.	Ever had sex

References

Abt Associates. 2016. *Design and Analysis Plan Quantitative Synthesis of Federally-Funded Teen Pregnancy Prevention Programs.* Report prepared for the Office of Adolescent Health. https://www.hhs.gov/ash/oah/sites/default/files/analysis-plan-050216-final-508.pdf

Borenstein, M., L. Hedges, J. Higgins, and H. Rothstein. 2009. *Introduction to Meta-Analysis.* West Sussex, UK: John Wiley and Sons.

Borman, G., G. Hewes, L. Overman, and S. Brown. 2003. "Comprehensive School Reform and Achievement: A Meta-Analysis." *Review of Educational Research* 73(2): 125-230.

Farb, A., and A. Margolis. 2016. "The Teen Pregnancy Prevention Program (2010-2015): Synthesis of Impact Findings." *American Journal of Public Health* (Supplement): s9–s15.

Jemmott, J. 1992. "Reductions in HIV risk-associated sexual behaviors among black male adolescents: Effects of an AIDS prevention intervention." *American Journal of Public Health* 82(3): 372-77.

Jemmott, J., L. Jemmott, G. Fong, and K. McCaffree. 1999. "Reducing HIV risk-associated sexual behavior among African American adolescents: Testing the generality of intervention effects." *American Journal of Community Psychology* 27(2): 161-187.

Juras, R., E. Tanner-Smith, M. Kelsey, M. Lipsey, and J. Layzer. 2019. "Adolescent Pregnancy Prevention: Meta-Analysis of Federally Funded Program Evaluations." *American Journal of Public Health* 109(4): e1-e8.

Kann, Laura, Tim McManus, William A. Harris, Shari L. Shanklin, Katherine H. Flint, ... and Kathleen A. Ethier. 2018. *Youth Risk Behavior Surveillance—United States, 2017* (Morbidity and Mortality Weekly Report, vol. 67, no. 8). https://www.cdc.gov/healthyyouth/data/yrbs/pdf/2017/ss6708.pdf

Knab, J., R. Cole, and S. Zief. 2016. "Challenges and Lessons Learned from Providing Large-Scale Evaluation Technical Assistance to Build the Adolescent Pregnancy Evidence Base." *American Journal of Public Health* (Supplement): s26–s28.

Lugo-Gil, J., A. Lee, D. Vohra, K. Adamek, J. Lacoe, and B. Goesling. 2016. *Updated Findings from the HHS Teen Pregnancy Prevention Evidence Review: July 2014 through August 2015.* http://tppevidencereview.aspe.hhs.gov/Reports.aspx

Petrosino, A., and H. Soydan. 2005. "The Impact of Program Developers as Evaluators on Criminal Recidivism: Results from Meta-Analyses of Experimental and Quasi-Experimental Research." *Journal of Experimental Criminology* 1(4) : 435-450.

Zief, S., J. Knab, and R. Cole. 2016. "A Framework for Evaluation Technical Assistance." *American Journal of Public Health* (Supplement): s24–s26.