



# ASPE

## Working Paper

### Programs to Reduce Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors: A Systematic Review

Brian Goesling, Silvie Colman, Christopher Trenholm  
Mathematica Policy Research

Mary Terzian, Kristin Moore  
Child Trends

**Purpose:** This paper presents findings from an ongoing systematic review of research on teen pregnancy and STI prevention programs, sponsored by the U.S. Department of Health and Human Services to help support evidence-based approaches to teen pregnancy prevention.

**Methods:** The review was conducted in four steps. First, multiple literature search strategies were used to identify relevant studies released from 1989 through roughly January 2011. Second, all studies identified through the literature search were screened against pre-specified inclusion criteria. Third, studies that met the inclusion criteria were assessed by teams of two trained reviewers for the quality and execution of their research designs. Fourth, for studies that passed this quality assessment, the review team extracted and analyzed information on the research design, study sample, evaluation setting, and program impacts.

**Results:** A total of 88 studies met the review criteria for study quality and were included in the data extraction and analysis. The studies examined a range of program models delivered in diverse settings. Most studies had mixed-gender and predominately African American research samples (70 percent and 51 percent, respectively). Randomized controlled trials accounted for the large majority (87 percent) of included studies. Most studies (76 percent) included multiple follow-ups, with sample sizes ranging from 62 to 5,244. Analysis of the study impact findings identified 31 program models with evidence of a statistically significant positive effect (and no adverse effects) on teen pregnancy, STIs, or sexual activity.

**Conclusions:** The number and rigor of evaluation studies on teen pregnancy and STI prevention programs has grown substantially since the late 1980s. Key strengths of the literature are the large number of randomized controlled trials, the common use of multiple follow-up periods, and attention to a broad range of program models delivered in diverse settings. Two main gaps are a lack of replication studies and the need for more research on Latino youth and other high-risk populations. In addressing these gaps, studies must overcome common limitations in research quality and reporting standards that have negatively affected prior research.

## **Introduction**

High rates of teen pregnancy, sexually transmitted infections (STIs), and associated sexual risk behaviors remain a troubling issue in the United States. Nationwide, 47 percent of high school students have had sexual intercourse, and 24 percent report having had four or more partners by graduation (CDC 2012). In 2011, nearly 40 percent of sexually active high school students had not used a condom during their last sexual intercourse (CDC 2012). These behaviors increase the risks of pregnancy and STIs, including HIV. Preliminary national data for 2011 indicate there were approximately 31.3 births per 1,000 females 15 to 19 years of age (Hamilton et al. 2012), a rate higher than in most other industrialized countries (United Nations 2012). In addition, estimates suggest that adolescents and young adults account for half of all new STI cases in the United States every year (Weinstock et al. 2004).

In response to these risks, in late 2009 and early 2010, the U.S. Congress authorized two new federal programs to support evidence-based approaches to teen pregnancy prevention: the Teen Pregnancy Prevention (TPP) program, administered by the Office of Adolescent Health (OAH) within the Office of the Assistant Secretary for Health (ASH), and the Personal Responsibility Education Program (PREP), authorized under the 2010 Patient Protection and Affordable Care Act (ACA) and administered by the Family and Youth Services Bureau (FYSB) within the Administration for Children and Families (ACF). The TPP program and PREP provide federal grant funding to local and state organizations to support evidence-based and promising approaches to teen pregnancy prevention. The programs are “tiered” so that the majority of funding goes toward replicating programs with existing evidence of effectiveness, while a smaller amount goes toward implementing and evaluating untested approaches. The first program grant announcements were released in spring 2010, for funding through 2015.

To help inform these new grant programs, in 2009, the U.S. Department of Health and Human Services (HHS) launched a systematic review of research on programs to reduce teen pregnancy, STIs, and associated sexual risk behaviors. The main purpose of the review is to identify, assess, and rate the rigor of program impact studies on teen pregnancy and STI prevention programs. The term “program” is defined broadly to include any structured intervention with an educational component, including classroom- or group-based curricula, individualized clinic services, and youth development approaches. The review findings are used to describe the strength of evidence supporting different program models and to identify programs with evidence of impacts on teen pregnancy, STIs, or associated risk behaviors. The review focuses only on assessing the research evidence. It does not examine the specific content, lesson plans, or delivery of any program.

The project has so far involved two rounds of reviews. An initial review of the evidence began in fall 2009, covering research released from 1989 through roughly December 2009. Summary findings were released in spring 2010 in conjunction with the TPP program and PREP Innovative Strategies (PREIS) grant announcements (OAH 2010). The findings were also highlighted in the 2010 state PREP grant announcement (ACF 2010). The review findings were later updated in spring 2012 to cover research published from roughly December 2009 through January 2011. Additional updates are planned for the future.

This paper presents additional findings from the review beyond those highlighted in the federal grant announcements. Specifically, the paper presents detailed counts of the number of studies that have been considered for review and the number that met the review criteria for study quality. The paper also presents information on the characteristics of included studies and the number of program models with evidence of impacts on teen pregnancy, STIs, or associated risk behaviors. Findings are used to draw conclusions about the relative strengths and weaknesses of the literature and directions for improving future research.

## Methods

The review is conducted for HHS by Mathematica Policy Research and a partner organization, Child Trends. HHS initiated and defined the scope of the review. Mathematica developed the review protocol at the direction of HHS. Mathematica also has lead responsibility for identifying studies, training and managing the review team, and synthesizing the results. The study assessments and data extraction are conducted by trained reviewers from Mathematica and Child Trends, all with doctorates or master's degrees in a relevant discipline. Reviewers are required to declare any conflicts of interest. None have been reported to date.

The review is conducted following a protocol developed by Mathematica at the direction of HHS (Mathematica Policy Research 2012). In developing the protocol, the review team drew on the standards and procedures used by established systematic reviews and evidence assessment groups (Advocates for Youth 2008; Blueprints for Violence Prevention 2011; Centers for Disease Control and Prevention 2011; Kirby 2007; Scher et al. 2006; What Works Clearinghouse 2012). The protocol was first developed in September 2009 for the initial review of the evidence. The protocol was later amended in December 2010 to expand the literature search through roughly January 2011.

### *Study Identification*

For the initial review of the evidence, studies were identified in four ways: (1) scanning the reference lists of prior systematic reviews and research syntheses (Advocates for Youth 2008; Ball and Moore 2008; Chin et al. 2012; Kim and Rector 2008; Kirby 2007; Oringanje et al. 2009; Scher et al. 2006); (2) searching the websites of relevant federal agencies and research or policy organizations; (3) issuing a public call for studies to identify new or unpublished research; and (4) having a research librarian conduct a keyword search of electronic citation databases. For the first update to the review findings, the review team also conducted a hand search of 10 relevant research journals and scanned the conference proceedings of five professional associations. The search covered both published and unpublished studies. Focusing only on published studies can lead to bias in systematic reviews (IOM 2011), since published studies tend to overrepresent favorable and statistically significant findings relative to null or negative findings. Additional details on the search strategy are available in the review protocol (Mathematica Policy Research 2012).

### *Study Screening*

All studies identified through the literature search were screened against prespecified inclusion criteria. To be eligible for review, a study had to examine the impacts of an intervention using quantitative data and statistical analysis and hypothesis testing. Both randomized controlled trials and quasi-experiments were eligible. Given this focus on studies designed to test causal program effects, the review excluded the large number of correlational studies that examine risk and protective factors for adolescent sexual activity, as well as qualitative studies of program implementation.

A study had to measure program impacts on a least one measure of pregnancy, STIs, or associated sexual risk behaviors (sexual initiation, frequency of sexual activity, recent sexual activity, number of sexual partners, or contraceptive use). Measures without established validity (for example, reports from males of their female partners' use of birth control pills) and composite scales of sexual risk behavior were excluded. The study sample had to consist of U.S. youth ages 19 or younger at the time of sample enrollment. International studies were excluded. The initial review of the evidence covered studies released from 1989 through roughly December 2009. The updated review findings added studies released from roughly December 2009 through January 2011.

The range of eligible program models included all programs eligible for funding under the federal TPP program. This included individualized and group curriculum-based programs, as well as youth development approaches. It excluded early childhood education programs, home visiting programs, high school dropout prevention programs, and broad state- or federal-policy changes.

The screening process had two steps. First, the review team screened the titles and abstracts of all studies identified through the literature search. Second, for studies that passed this first stage of screening, the review team obtained full text of the studies for another round of screening. In some cases, study authors reported findings from a single study in multiple journal articles or reports (for example, Basen-Engquist et al. 2001; Coyle et al. 1999, 2001; Kirby et al. 2004, 2011). To avoid double-counting these studies, the review team linked together any articles or reports generated from the same underlying study. In other cases, a single journal article or report presented findings from multiple studies or a multi-armed trial (for example, Jemmott et al. 1998). In these cases, the review team counted each study or trial arm separately.

### *Study Quality Assessment*

All studies that met the review inclusion criteria were assessed by teams of two trained reviewers for the quality and execution of their research designs. The reviewers made their assessments following a prespecified set of standards documented in the review protocol. Differences of opinion were resolved through consensus.

At the end of the assessment, each study was assigned a quality rating of *high*, *moderate*, or *low* according to the risk of bias in the study's impact estimates. The highest quality rating was reserved for randomized controlled trials with low attrition of sample members, no reassignment

of sample members across conditions, and no systematic differences in the timing or mode of data collection across the treatment and control groups. Cluster randomized trials were required to have at least two clusters (of schools, classrooms, and so on) assigned to each condition. Studies that received this highest rating were considered to provide the most credible estimates of program effects.

The moderate quality rating was considered for (1) quasi-experimental comparison group designs and (2) randomized controlled trials that did not meet all the review criteria for the highest quality rating. To receive a moderate rating, a study had to demonstrate baseline equivalence of the research groups (that is, the program and comparison groups) on three key demographic characteristics: age, gender, and race/ethnicity. For studies with sample members at least 14 years old at baseline, the study authors also had to demonstrate evidence of baseline equivalence for at least one outcome measure. This criterion was not applied to studies with younger sample members who were less than 14 years old, because rates of sexual risk behaviors are very low for this age group. As required for the highest study rating, the timing and mode of data collection had to be the same across program and comparison groups, and cluster designs had to have at least two clusters (of schools, classrooms, and so on) in each group.

The lowest quality rating was applied to studies that did not meet the review standards for either a high or moderate rating. Low-rated studies were excluded from the subsequent data extraction and analysis, as the risk of bias in these studies was considered too high to yield credible estimates of program effects. A more detailed description and justification of the study ratings is presented in the review protocol (Mathematica Policy Research 2012).

### *Data Extraction*

For studies that met the review criteria for a high or moderate quality rating, the review team extracted information on the program model tested, evaluation setting, study sample, and research design. The review team also extracted detailed information on the program impact estimates. Specifically, for each relevant outcome, the review team recorded the name and description of the outcome measure, length of follow-up, analytic sample used to estimate the program impact, magnitude of the impact estimate, reported statistical confidence interval or associated standard error of the estimate, reported *p*-value or other associated test statistic, and statistical significance level as reported by the study authors.

### *Analysis*

The analysis had three steps. First, all studies included in the data extraction were tabulated by program type and other key features of the evaluation setting, study sample, and research design. These tabulations were used to summarize and describe characteristics of the evidence base. Second, the program impact findings were used to identify program models with evidence of a statistically significant positive effect (and no adverse effects) on at least one of the following outcomes: sexual activity (initiation, recent sexual activity, frequency of sexual activity, or number of partners); contraceptive use or consistency of use; STIs; or pregnancy or birth. To reduce the possibility of detecting chance findings due to multiple hypothesis testing, this assessment was limited to program impacts estimated for either the full study sample or a subgroup defined by gender or baseline sexual experience. Third, the review team tabulated the

resulting program models according to the strength and nature of the supporting evidence (for example, the duration of program impacts, and whether the observed impacts have been replicated by two or more studies).

The review did not consider evidence from subgroups defined by sexual activity at follow-up. To estimate program impacts on measures such as condom or contraceptive use, studies often limit their analytic samples to only those youth who report being sexually active at follow-up. These impact estimates are at risk of bias, however, because the size and composition of this “endogenous” subgroup of sexually active youth may be affected by the intervention (Colman 2012). To minimize the risk of bias, the review focused only on analytic samples defined by characteristics independent of the intervention—namely, either the full study sample or subgroups defined by gender or sexual experience at baseline.

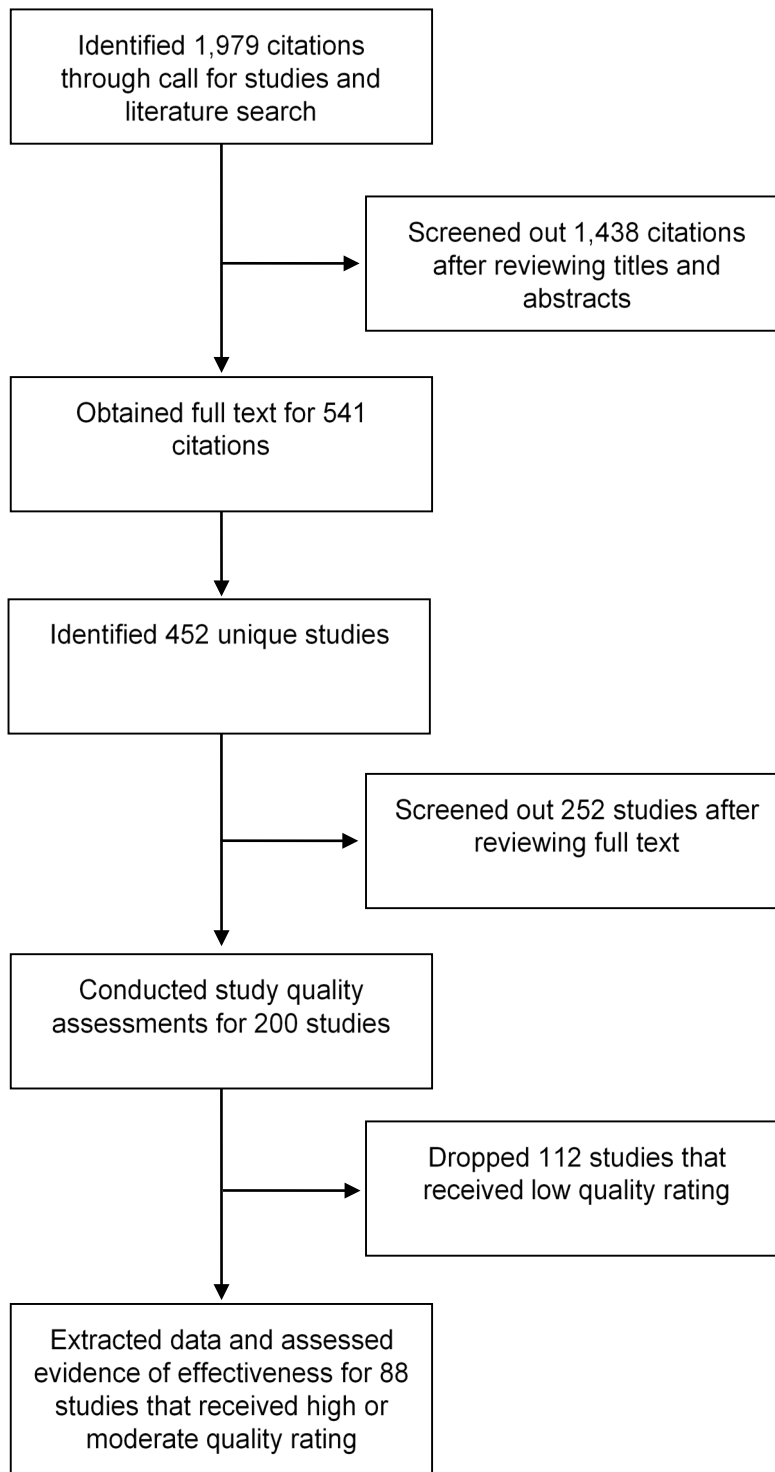
The review did not require conducting a quantitative meta-analysis. The main purpose of the review was to describe the evidence base and identify program models with evidence of positive impacts on youth outcomes. In identifying these programs, there was little need or opportunity to average impact findings for the same program across multiple studies, because the large majority of teen pregnancy and STI prevention programs have been evaluated only once (shown below). Unlike some prior systematic reviews of the teen pregnancy prevention literature (Chin et al. 2012; Gavin et al. 2010; Oringanje et al. 2009; Scher et al. 2006), this review did not aim to summarize or compare findings across broader groups of program models.

## Results

More than 1,900 citations were found through the literature search (Figure 1). From this initial citation list, 1,438 (73 percent) were screened out by title and abstract. Full text articles were obtained for 541 citations, and from these citations, the review team identified 452 unique studies. An additional 252 studies were screened out after reviewing the full text, and 112 studies were dropped for failing to meet the review criteria for a high or moderate study quality rating. A total of 88 studies met the review criteria for a high or moderate rating and were included in the final data extraction and analysis.

The studies tested a mix of program models delivered in diverse settings (Table 1). Nearly half the included studies (47 percent) examined impacts for sexuality education programs—defined broadly as curriculum-based programs providing general information on teen pregnancy and STI prevention, including the use of contraceptives. Other studies examined abstinence-based programs (19 percent), clinic-based programs offering individualized services (11 percent), youth development programs (11 percent), or programs for specialized populations such as pregnant or parenting teens or youth in the foster care or juvenile justice systems (11 percent). Most programs were delivered in after school or community-based organizations (38 percent) or in school during the regular school day (29 percent). Among the in-school programs, more were delivered in middle schools than in high schools or elementary schools. Relatively few studies were set in community health clinics (16 percent) or specialized settings such as juvenile justice facilities (11 percent). Six percent of the studies combined data across multiple settings.

**Figure 1. Flow of Citations and Studies Through the Review**



**Table 1. Program and Sample Characteristics of Included Studies (n = 88)**

Characteristic	Number of Studies (Percentage)
Program type	
Abstinence-based	17 (19)
Clinic-based	10 (11)
Sexuality education	41 (47)
Programs for special populations <sup>a</sup>	10 (11)
Youth development	10 (11)
Program length	
Fewer than 10 sessions	57 (65)
10 to 20 sessions	13 (15)
More than 20 sessions	18 (20)
Evaluation setting	
After school/community-based	33 (38)
Health clinic	14 (16)
In-school	26 (29)
Elementary school	2 (2)
Middle school	18 (20)
High school	6 (7)
Multiple settings	5 (6)
Specialized setting <sup>b</sup>	10 (11)
Average age group	
13 or younger	39 (44)
14 to 17	39 (44)
18 or 19	10 (11)
Majority racial/ethnic group	
African American	45 (51)
Asian	1 (1)
Latino	17 (19)
White	25 (28)
Sex	
Both sexes	62 (70)
Female only	19 (22)
Male only	7 (8)

<sup>a</sup> Comprises programs designed specifically for use with youth in the juvenile justice system, foster care youth, homeless/runaway youth, pregnant or parenting teens, and other specialized populations.

<sup>b</sup> Comprises juvenile justice facilities, residential facilities for substance dependent youth, and other specialized settings



Most studies (62 percent) included mixed-sex samples, and slightly more than half featured predominately African American samples (51 percent). Among the single-sex studies, more focused on females than on males. The distribution of studies by age group found equal proportions of studies with youth ages 14 to 17 (44 percent) and youth ages 13 and under (44 percent). Fewer studies focused on older youth ages 18 or 19 (11 percent).

Randomized controlled trials accounted for a large majority of included studies (Table 2). The most common method was random assignment of individual youth (44 percent), but nearly an equal proportion (41 percent) assigned youth in clusters such as schools. Two studies used a mix of individual- and cluster-based assignment. Few studies used quasi-experimental designs (13 percent). Sample sizes ranged from a low of 62 to a high of 5,244, with a median sample size of 447.

Most studies involved more than one follow-up period. Indeed, nearly half (47 percent) involved three or more follow-ups. In most studies, the first follow-up was conducted immediately after the intervention. The length of the last follow-up ranged from immediately after the intervention to 15 years after the intervention ended. The most common outcome measures examined were behavioral: sexual activity (86 percent) and contraceptive use and/or consistency (80 percent). Fewer studies examined impacts on STIs (23 percent) or pregnancy or births (28 percent).

Among the 88 studies included in the analysis, the review team identified 78 unique program models (Figure 2). The number of unique program models is only slightly lower than the number of studies because few programs have been evaluated more than once. Among the 78 unique programs, 34 had null findings: the review team found no evidence of a statistically significant positive impact on teen pregnancy, STIs, or sexual risk behaviors for either the full sample or a key subgroup. Another 13 programs were dropped because evidence of positive impacts was shown only for an endogenous subgroup defined by sexual activity at follow-up. For the remaining 31 programs, the review team found evidence of a statistically significant positive impact (and no adverse effects) on at least one targeted outcome measure.

Most of the positive impact evidence focused on measures of sexual activity and contraceptive use (Table 3). Among the 31 program models with evidence of effectiveness, 22 had impacts on a measure of sexual activity, 14 had impacts on a measure of contraceptive use or consistency, 5 had impacts on STIs, and 5 had impacts on pregnancy or birth. Of the 31 programs, 19 had impacts on only one of these four categories of outcome measures, 10 had impacts on two or three categories of outcomes, and one program (DiClemente et al. 2004) had impacts on all four categories of outcome measures.

One program model (Jemmott 1992; Jemmott et al. 1999, 2010) had evidence of impacts replicated across multiple studies (Table 4). The other programs were supported by evidence from a single qualifying impact study. In most cases, the supporting impact study was a randomized controlled trial that met the review criteria for a high study quality rating. The remaining programs were supported by either a quasi-experimental study (four programs) or a randomized trial that met the review criteria for a moderate (but not high) study quality rating (six programs). A total of 9 programs were supported by evidence of subgroup impacts, defined

**Table 2. Design Characteristics of Included Studies (n = 88)**

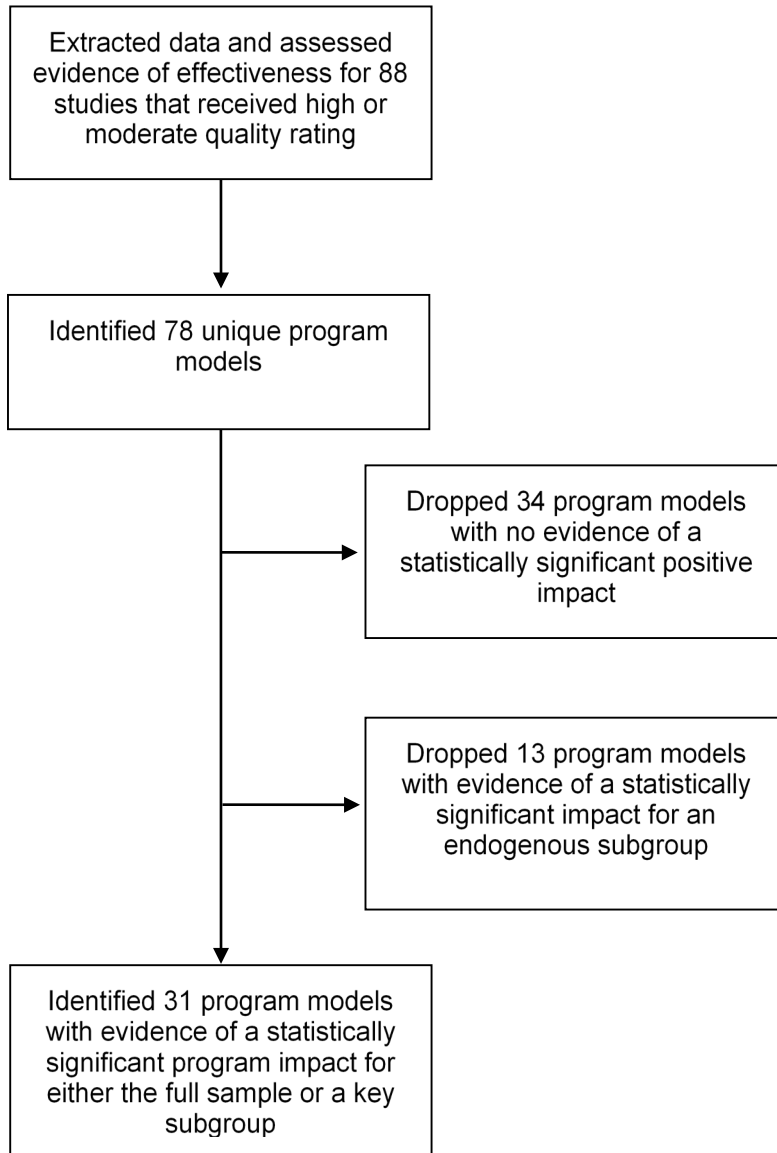
<b>Characteristic</b>	<b>Number of Studies (Percentage)</b>
Study Design	
Randomized controlled trial	77 (87)
Cluster	36 (41)
Individual	39 (44)
Mixed <sup>a</sup>	2 (2)
Quasi-experimental design	11 (13)
Sample size	
Smallest	n = 62
Median	n = 447
Largest	n = 5,244
Number of follow-up surveys	
One	21 (24)
Two	25 (28)
Three	24 (27)
Four or more	18 (20)
Length of first follow-up <sup>b</sup>	
Shortest	0 months
Median	0 months
Longest	72 months
Length of last follow up <sup>b</sup>	
Shortest	0 months
Median	12 months
Longest	180 months
Outcome measures <sup>c</sup>	
Sexual activity	76 (86)
Sexual initiation/abstinence	44 (50)
Recent sexual activity	39 (44)
Number of sexual partners	43 (49)
Frequency of sexual activity	24 (27)
Contraceptive use and/or consistency	70 (80)
Sexually transmitted infections (STIs)	20 (23)
Pregnancy or birth	25 (28)

<sup>a</sup> Some participants were randomly assigned in clusters and others as individuals.

<sup>b</sup> Measured as months since the end of the intervention.

<sup>c</sup> Percentages do not sum to 100 because some studies measure more than one outcome.

**Figure 2. Identifying Program Models with Evidence of Effectiveness**



**Table 3. Programs with Evidence of Favorable Impacts (n = 31)**

<b>Program</b>	<b>Sexual Activity</b>	<b>Contraceptive Use</b>	<b>STIs</b>	<b>Pregnancy or Birth</b>
Aban Aya Youth Project	+	na	na	na
Adult Identity Mentoring (Project AIM)*	+	na	na	na
All4You!	+	+	na	na
Assisting in Rehabilitating Kids (ARK)*	+	+	na	na
Be Proud! Be Responsible!*	+	+	na	na
Be Proud! Be Responsible! Be Protective!*	+	o	na	na
Becoming a Responsible Teen (BART)*	+	+	na	na
Children's Aid Society (CAS)—Carrera Program*	+	na	na	+
¡Cuidate!*	+	+	na	na
Draw the Line/Respect the Line*	+	na	na	na
FOCUS*	+	o	na	na
Heritage Keepers Abstinence Education	+	na	na	na
Horizons*	na	+	+	na
It's Your Game: Keep it Real	+	na	na	na
Making a Difference!*	+	o	na	na
Making Proud Choices!*	o	+	na	na
Project TALC*	o	na	na	+
Promoting Health Among Teens! Abstinence Only Intervention*	+	o	na	na
Promoting Health Among Teens! Comprehensive Abstinence and Safer Sex Intervention*	+	o	na	na
Reducing the Risk	o	+	na	o
Rikers Health Advocacy Program (RHAP)	o	+	na	na
Raising Healthy Children	+	o	+	+
Respeto/Proteger*	na	+	na	na
Safer Choices*	o	+	na	na
Safer Sex	+	o	na	na
Sexual Health and Adolescent Risk Prevention (SHARP)	na	+	na	na
SiHLE*	+	+	+	+
Sisters Saving Sisters*	+	+	+	na
Teen Health Project	+	na	na	na
Teen Outreach Program*	o	na	na	+
What Could You Do?*	+	o	+	na

\* Denotes programs supported by a randomized controlled trial that met the review criteria for a high rating.

+ = statistically significant program impact; o = no statistically significant program impact; na = not available (either not measured or did not meet review criteria).

by either gender or sexual experience at baseline. The other 22 programs had evidence of full sample impacts. Fewer than half of the programs found evidence of sustained impacts for more than 12 months post intervention.

## **Discussion**

The number and rigor of evaluation studies on teen pregnancy and STI prevention programs have grown substantially since the 1980s. To help summarize this research, a systematic review of the literature, sponsored by HHS as part of the federal government's recent efforts to promote evidence-based approaches to teen pregnancy prevention, identified and assessed some 200 program impact studies released from 1989 through early 2011. Of the studies assessed, a total of 88 met the review criteria for study design and execution. Analysis of the study impact findings identified 31 program models with evidence of a statistically significant positive impact on teen pregnancy, STIs, or associated sexual risk behaviors.

### *Prominence of Randomized Trials*

A main strength of the literature is the large number of randomized controlled trials. In some areas of program evaluation and policy research, randomized controlled trials either are not feasible or are considered an unrealistic standard. For example, in studies of community-level interventions, there are often too few communities involved to support a sufficiently powered randomized trial. By contrast, in this review of the teen pregnancy and STI prevention literature, a large majority of included studies (87 percent) used randomized designs. These designs have been used successfully with all types of program models and in diverse settings, ranging from schools (such as Coyle et al. 1999, 2001; Tortolero et al. 2009) to juvenile justice facilities (such as Bryan et al. 2009; Magura et al. 1994). These findings strongly suggest that randomized controlled trials are a realistic expectation for the teen pregnancy and STI prevention literature and the foundation on which future research should be built.

### *Use of Multiple Follow-Up Surveys*

Another strength of the literature is the common use of multiple follow-up surveys. Conducting multiple follow-ups enables researchers to test both short- and longer-term program impacts as well as the mechanisms or pathways through which programs work. For example, studies often use shorter-term follow-up surveys to measure program impacts on key mediating outcomes such as skills, attitudes, and intentions. Longer-term follow-ups are often better for measuring program impacts on behaviors or health outcomes, which can take longer to emerge. Researchers are clearly attuned to the benefits of multiple follow-up surveys, as just over three-quarters (76 percent) of the studies included in the review featured more than one follow-up.

Future research should more carefully consider the best timing for follow-up surveys. A relatively common approach is to conduct follow-ups at regular six-month intervals—for example, immediately after the intervention and again 6 and 12 months later. For most studies, this schedule is sufficient for measuring program impacts on sexual activity or contraceptive use, but it does not always allow for measuring impacts on pregnancy and STIs, two outcomes that take longer to unfold. The frequent use of a 6-month follow-up schedule is likely one reason studies tend to focus more on behavioral outcomes such as sexual activity and contraceptive use

than on the longer-term health outcomes of STIs and pregnancy. Future studies should consider whether alternative follow-up schedules—for example, two follow-ups conducted 6- and 18-months after the intervention—allow for testing impacts on a broader range of outcomes at similar cost.

Less clear is the ideal number of follow-ups. A full 20 percent of studies included in the review featured four or more follow-ups. However, unless the program is especially long or the study is designed to test very long-term effects (Hawkins et al. 1999, 2008; Lonczak et al. 2002), the analytic value of four or more follow-ups is often modest, at increased cost. Adding a fourth or fifth follow-up can also lead to overly optimistic assessments of program effectiveness—namely, if analyses using each round of follow-up data are treated as independent assessments. To help avoid chance findings, researchers must apply appropriate statistical adjustments for multiple hypothesis testing to control the type-1 error rate. Given these concerns, resources are often better spent increasing the study sample size or obtaining higher response rates than adding a fourth or fifth follow-up.

### *Diversity of Program Models and Settings*

The literature is also strong in its focus on a broad range of program models delivered in diverse settings. High-quality studies have been conducted with programs ranging from curriculum-based abstinence and sexuality education programs to individualized clinic-based services, in settings ranging from schools to residential substance abuse and mental health facilities.

This diversity is important for two reasons. For one, there is no single recipe for success in improving adolescent sexual health outcomes. Prior studies have shown that, among similar types of programs (for example, among all clinic-based or youth development programs), there is often significant variation in program impacts (Chin et al. 2012; Kirby 2007; Underhill et al. 2007a, 2007b). Some programs have demonstrated evidence of success whereas others have not. Even in cases in which two programs have tried following the same approach, the impacts on youth outcomes have differed (see for example, Philliber et al. 2002; Kirby et al. 2005). Until there is more rigorous evidence about why some program models are more effective than others, the field is best served by continuing to test a range of programmatic approaches.

Diversity is also important to meet the unique needs and interests of local communities. No single program model is right for every population and setting. For example, schools have different programmatic needs than community-based organizations or institutional settings such as juvenile justice facilities. Similarly, youth in rural areas may respond differently to programs originally developed in urban settings (Stanton et al. 2005, 2006). Because no one size fits all, it is important to have a variety program models available for implementation, and this in turn requires a research literature that is equally broad in focus.

### *Need for Studies of Replication and Scale-Up*

The biggest gap in the literature is a lack of replication studies. Researchers increasingly recognize replication as a key step in the process of identifying effective interventions (Valentine et al. 2011). Yet the findings from this review show that only one program model (Jemmott 1992; Jemmott et al. 1999, 2010) in the teen pregnancy and STI prevention literature has

demonstrated evidence of program impacts across multiple studies. Instead, most of the current evidence base consists of single small-scale “efficacy trials” (Flay et al. 2005) conducted in closely managed settings, often by the program developers. These efficacy studies are important for establishing initial evidence of program impacts, but to determine whether the impacts generalize to broader populations and more real-world conditions, they must be followed by subsequent effectiveness or replication studies, ideally conducted independently of the program developer.

Research from outside the field of teen pregnancy and STI prevention finds that efficacy trials typically produce larger impacts than when programs are “scaled up” as in effectiveness or replication studies. For example, a recent review article of early intervention programs for crime and delinquency prevention suggests that program impacts may be “discounted” by up to 50 percent when programs are implemented on a very large scale (Welsh et al. 2010). For the literature on teen pregnancy and STI prevention programs, these findings give warning that the existing evidence in support of some program models may weaken as the research literature expands.

#### *Need for Research on High-Risk Populations*

There is also a need for more research on specific populations at high risk for teen pregnancies or STIs, most notably the large and growing U.S. Latino population. Latino youth currently have the highest teen birth rate of all major racial/ethnic groups in the United States—49.4 births per 1,000 teens ages 15 to 19, compared with the national average of 31.3 per 1,000 teens (Hamilton et al. 2012). Yet this review found that less than a quarter of the relevant impact studies conducted to date had predominately Latino research samples. Only two program models designed specifically for use with Latino youth have found evidence of a statistically significant program impact on sexual risk behaviors (Lesser et al. 2009; Villarruel et al. 2006), and one of these programs is used only with teen parents. Five other programs with evidence of effectiveness were evaluated with predominately Latino research samples (Coyle et al. 2004; Coyle et al. 2006; Koniak-Griffin et al. 2003; Rotheram-Borus et al. 2003, 2004; Tortolero et al. 2009), but the programs themselves were not culturally tailored specifically for Latinos.

There is also little evidence for smaller but high-risk groups such as pregnant or parenting teens, youth living in foster care, and American Indian and Alaska Natives. Only two studies included in the review focused specifically on pregnant and parenting teens (Koniak-Griffin et al. 2003; Lesser et al. 2009), one examined youth in foster care (Kerr et al. 2009), and none focused specifically on American Indians and Alaska Natives. All of these groups are at above-average risk of teen pregnancy (Bilaver and Courtney 2006; Klerman 2004; Suellentrop and Hunter 2009) and thus merit the attention of future research.

#### *Need for Improved Research Quality and Reporting*

In addressing these gaps, studies also must strive for improved research quality and reporting standards. More than half the studies considered for this review did not pass the bar for study design and execution. Three common problems that led to a downgrade in study rating were high rates of sample attrition in randomized trials, poorly matched comparison groups in quasi-experimental studies, and the use of a cluster design with only one cluster assigned to each

research group. Some studies failed to report a complete description of the study design and execution. The median sample size was 447, which may be too small to detect substantively meaningful program effects (Kirby 2007).

Other common problems that did not factor directly into our review but represent key areas for improvement were failure to properly adjust statistical significance tests for multiple hypothesis testing or the use of a clustered study design, insufficient reporting of consent rates and the timing of consent in randomized trials, and a heavy reliance on subgroup estimates to demonstrate evidence of program effects (Donner and Klar 2004; Schochet 2009). Studies also failed to consistently report effect size information. More than one-third of all statistically significant impact estimates were lacking necessary information to calculate an effect size. All of these issues can be addressed through more careful execution of the evaluation study and more complete reporting of future research studies.

### *Summary and Conclusions*

Like any field of research, the evidence base on teen pregnancy and STI prevention programs has both strengths and weaknesses. Key strengths are the large number of randomized controlled trials, the common use of multiple follow-up surveys, and attention to a broad range of program models delivered in diverse settings. Main weaknesses are the current lack of replication studies, the need for more research on Latinos and other high-risk populations, and a general need for improved research quality and reporting.

Addressing these gaps is important not only to advance the research literature but also to inform future policy initiatives and programming decisions. Efforts like the new federal TPP program and PREP reflect a growing emphasis on the use of research evidence in policy and programming decisions (Haskins and Baron 2011). Both within and outside the federal government, funding decisions are increasingly conditioned on available evidence of program effectiveness. Such reliance on evidence has the potential to maximize the effectiveness of new policy initiatives and make the best use of scarce public resources. But evidence-based policies and programs are only as strong as the supporting research evidence. In addition to the need for rigorous program impact studies, there is also a need for research on program implementation, measures of fidelity, identification of core components, and acceptable adaptations. Such a strong and expanding research literature is necessary to ensure that evidence-based approaches to teen pregnancy prevention make a positive difference in the lives of youth.



## References

- Administration for Children and Families. “State Personal Responsibility Education Program.” Washington, DC: Administration on Children, Youth, and Families (ACYF)/Family and Youth Services Bureau (FYSB), Administration for Children and Families, U.S. Department of Health and Human Services, 2010.
- Advocates for Youth. “Science and Success, 2nd edition.” Washington, DC: Advocates for Youth, 2008.
- Ball, V., and K.A. Moore. “What Works for Adolescent Reproductive Health: Lessons from Experimental Evaluations of Programs and Interventions.” Washington, DC: Child Trends, 2008.
- Basen-Engquist, K., K. Coyle, G. Parcel, D. Kirby, S. Banspach, and S. Carvajal. “Schoolwide Effects of a Multicomponent HIV, STD, and Pregnancy Prevention Program for High School Students.” *Health Education & Behavior*, vol. 28, no. 2, 2001, pp. 166–185.
- Bilaver, L.A., and M.E. Courtney. “Foster Care Youth.” Washington, DC: National Campaign to Prevent Teen Pregnancy, 2006.
- Blueprints for Violence Prevention Selection Criteria. Available at: [<http://www.colorado.edu/cspv/blueprints/criteria.html>]. Accessed August 11, 2012.
- Bryan, A. D., S. J. Schmiege, and M. R. Broaddus. “HIV Risk Reduction Among Detained Adolescents: A Randomized, Controlled Trial.” *Pediatrics*, vol. 124, no. 6, 2009, pp. e1180–e1188.
- Centers for Disease Control and Prevention. “HIV Risk Reduction Efficacy Review Methods.” Available at: [[https://www.cdc.gov/hiv/pdf/research/interventionresearch/compendium/rr/cdc-hiv-rr\\_review\\_methods.pdf](https://www.cdc.gov/hiv/pdf/research/interventionresearch/compendium/rr/cdc-hiv-rr_review_methods.pdf)]. Accessed August 2, 2016.
- Centers for Disease Control and Prevention. “Youth Risk Behavior Surveillance—United States, 2011.” *MMWR*, vol. 61, no. SS-4, 2012, pp. 1–162.
- Chin, H.B. T. A. Sipe, R. Elder, et al. “The Effectiveness of Group-Based Comprehensive Risk-Reduction and Abstinence Education Interventions to Prevent or Reduce the Risk of Adolescent Pregnancy, Human Immunodeficiency Virus, and Sexually Transmitted Infections: Two Systematic Reviews for the Guide to Community Preventive Services.” *American Journal of Preventive Medicine*, vol. 42, 2012, pp. 272–294.
- Colman, S. “Estimating Program Impacts for a Subgroup Defined by Post-Intervention Behavior: Why is it a Problem? What is the Solution?” Evaluation Technical Assistance Brief for OAH & ACYF Teenage Pregnancy Prevention Grantees. Princeton, NJ: Mathematica Policy Research, December 2012.

Coyle, K. K., D. B. Kirby, B. V. Marin, C. A. Gomez, and S. E. Gregorich. "Draw the Line/Respect the Line: A Randomized Trial of a Middle School Intervention to Reduce Sexual Risk Behaviors." *American Journal of Public Health*, vol. 94, 2004, pp. 843–851.

Coyle, K. K., D. B. Kirby, L. E. Robin, S. W. Banspach, E. Baumler, and J. R. Glassman. "All4You! A Randomized Trial of an HIV, other STDs, and Pregnancy Prevention Intervention for Alternative School Students." *AIDS Education and Prevention*, vol. 18, no. 3, 2006, pp. 187–203.

Coyle, K., K. Basen-Engquist, D. Kirby, G. Parcel, S. Banspach, J. Collins, et al. "Safer Choices: Reducing Teen Pregnancy, HIV, and STDs." *Public Health Reports*, vol. 116, no. 6, 2001, pp. 82–93.

Coyle, K., K. Basen-Engquist, D. Kirby, G. Parcel, S. Banspach, R. Harrist, et al. "Short-Term Impact of Safer Choices: A Multicomponent, School-Based HIV, Other STD, and Pregnancy Prevention Program." *Journal of School Health*, vol. 69, no. 5, 1999, pp. 181–188.

DiClemente, R. J., G. M. Wingood, K. F. Harrington, D. L. Lang, S. L. Davies, E. W. Hook, et al. "Efficacy of an HIV Prevention Intervention for African American Adolescent Girls: A Randomized Controlled Trial." *JAMA: Journal of the American Medical Association*, vol. 292, no. 2, 2004, pp. 171–179.

Donner, A., and N. Klar. "Pitfalls of and Controversies in Cluster Randomized Trials." *American Journal of Public Health*, vol. 94, no. 3, 2004, pp. 416–422.

Flay, B. R., A. Biglan, R. F. Boruch, F. González Castro, D. Gottfredson, S. Kellam, E. K. Mościcki, S. Schinke, J. C. Valentine, and P. Ji. "Standards of Evidence: Criteria for Efficacy, Effectiveness, and Dissemination." *Prevention Science*, vol. 6, no. 3, 2005, pp. 151–175.

Gavin, E. G., R. F. Catalano, C. David-Ferdon, et al. "A Review of Positive Youth Development Programs That Promote Adolescent Sexual and Reproductive Health." *Journal of Adolescent Health*, vol. 46, 2010, pp. S75–S91.

Haskins, R. and J. Baron. "Building the Connection Between Policy and Evidence: The Obama Evidence-Based Initiatives." London, UK: NESTA, September 2011.

Hamilton, B. E., J. A. Martin, and S. J. Ventura. "Births: Preliminary Data for 2011." *National Vital Statistics Reports* web release, vol. 61, no. 5. Hyattsville, MD: National Center for Health Statistics, October 2012.

Hawkins, J. D., R. F. Catalano, R. Kosterman, R. Abbott, and K. G. Hill. "Preventing Adolescent Health-Risk Behaviors by Strengthening Protection During Childhood." *Archives of Pediatrics & Adolescent Medicine*, vol. 153, no. 3, 1999, pp. 226–234.

Hawkins, J. D., R. Kosterman, R. F. Catalano, K. G. Hill, and R. D. Abbott. "Effects of Social Development Intervention in Childhood 15 Years Later." *Archives of Pediatrics & Adolescent Medicine*, vol. 162, no. 12, 2008, pp. 1133–1141.

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

Jemmott, J. B. III. “Reductions in HIV Risk-Associated Sexual Behaviors Among Black Male Adolescents: Effects of an AIDS Prevention Intervention.” *American Journal of Public Health*, vol. 82, no. 3, 1992, pp. 372–377.

Jemmott, J. B., L. S. Jemmott, and G. T. Fong. “Abstinence and Safer Sex HIV Risk-Reduction Interventions for African American Adolescents: A Randomized Controlled Trial.” *JAMA: Journal of the American Medical Association*, vol. 279, no. 19, 1998, pp. 1529–1536.

Jemmott, J. B., L. S. Jemmott, G. T. Fong, and K. McCaffree. “Reducing HIV Risk-Associated Sexual Behavior Among African American Adolescents: Testing the Generality of Intervention Effects.” *American Journal of Community Psychology*, vol. 27, no. 2, 1999, pp. 161–187.

Jemmott, J. B. III, L. S. Jemmott, G. T. Fong, and K. H. Morales. “Effectiveness of an HIV/STD Risk-Reduction Intervention for Adolescents When Implemented by Community-Based Organizations: A Cluster-Randomized Controlled Trial.” *American Journal of Public Health*, vol. 100, no. 4, 2010, pp. 720–726.

Kerr, D. C. R., L. D. Leve, and P. Chamberlain. “Pregnancy Rates Among Juvenile Justice Girls in Two Randomized Controlled Trials of Multidimensional Treatment Foster Care.” *Journal of Counseling and Clinical Psychology*, vol. 77, no. 3, 2009, pp. 588–593.

Kim, C. C., and R. Rector. “Abstinence Education: Assessing the Evidence.” Washington, DC: The Heritage Foundation, 2008.

Kirby, D. “Emerging Answers 2007: Research Findings on Programs to Reduce Teen Pregnancy and Sexually Transmitted Diseases.” Washington, DC: National Campaign to Prevent Teen and Unplanned Pregnancy, 2007.

Kirby, D. B., E. Baumler, and K. K. Coyle. “The Impact of ‘Safer Choices’ on Condom Use and Contraceptive Use among Sexually Experienced Students at Baseline.” Unpublished manuscript, 2011.

Kirby, D. B., E. Baumler, K. K. Coyle, K. Basen-Engquist, G. S. Parcel, R. Harrist, et al. “The ‘Safer Choices’ Intervention: Its Impact on the Sexual Behaviors of Different Subgroups of High School Students.” *Journal of Adolescent Health*, vol. 35, no. 6, 2004, pp. 442–452.

Kirby, D., T. Rhodes, and S. Campe. “Implementation of Multi-Component Youth Programs to Prevent Teen Pregnancy Modeled After the Children's Aid Society-Carrera Program.” Scotts Valley, CA: ETR Associates, 2005.

Klerman, Lorraine V. “Another Chance: Preventing Additional Births to Teen Mothers.” Washington, DC: National Campaign to Prevent Teen Pregnancy, 2004.

Koniak-Griffin, D., J. Lesser, A. Nyamathi, G. Uman, J. A. Stein, and W. G. Cumberland. "Project CHARM: An HIV Prevention Program for Adolescent Mothers." *Family and Community Health*, vol. 26, 2003, pp. 94–107.

Lesser, J., D. Koniak-Griffin, R. Huang, S. Takayanagi, and W. G. Cumberland. "Parental Protectiveness and Unprotected Sexual Activity Among Latino Adolescent Mothers and Fathers." *AIDS Education & Prevention*, vol. 21, 2009, pp. 88–102.

Lonczak, H. S., R. D. Abbott, J. D. Hawkins, R. Kosterman, and C. F. Catalano. "Effects of the Seattle Social Development Project on Sexual Behavior, Pregnancy, Birth, and Sexually Transmitted Disease Outcomes by Age 21 Years." *Archives of Pediatrics & Adolescent Medicine*, vol. 156, no. 5, 2002, pp. 438–447.

Magura, S., S. Y. Kang, and J. L. Shapiro. "Outcomes of Intensive AIDS Education for Male Adolescent Drug Users in Jail." *Journal of Adolescent Health*, vol. 15, no. 6, 1994, pp. 457–463.

Mathematica Policy Research. "Identifying Programs That Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors: Review Protocol, Version 2.0." Available at: [[http://tppevidencereview.aspe.hhs.gov/pdfs/PPRER\\_Protocol\\_v2\\_4.19.pdf](http://tppevidencereview.aspe.hhs.gov/pdfs/PPRER_Protocol_v2_4.19.pdf)]. Accessed August 2, 2016.

Office of Adolescent Health. "Teenage Pregnancy Prevention: Replication of Evidence-based Programs." Washington, DC: Office of Adolescent Health, Office of Public Health and Science, U.S. Department of Health and Human Services, April 2010.

Oringanje, C., M. M. Meremikwu, H. Eko, et al. "Interventions for Preventing Unintended Pregnancies Among Adolescents." *Cochrane Database of Systematic Reviews*, no. 4, 2009.

Philliber, S., J. W. Kaye, S. Herrling, and E. West. "Preventing Pregnancy and Improving Health Care Access Among Teenagers: An Evaluation of the Children's Aid Society-Carrera Program." *Perspectives on Sexual and Reproductive Health*, vol. 34, no. 5, 2002, pp. 244–251.

Rotheram-Borus, M. J., M. Lee, N. Leonard, Y. Y. Lin, L. Franzke, E. Turner, M. Lightfoot, and M. Gwadz. "Four-Year Behavioral Outcomes of an Intervention for Parents Living with HIV and Their Adolescent Children." *AIDS*, vol. 17, 2003, pp. 1217–1225.

Rotheram-Borus, M. J., M. Lee, Y. Y. Lin, and P. Lester. "Six-Year Intervention Outcomes For Adolescent Children of Parents with the Human Immunodeficiency Virus." *Archives of Pediatric & Adolescent Medicine*, vol. 158, 2004, pp. 742–748.

Scher, L., R. A. Maynard, and M. Stagner M. "Interventions Intended to Reduce Pregnancy-Related Outcomes Among Adolescents." *Campbell Systematic Reviews*, vol. 12, 2006.

Schochet, P. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Education Review*, vol. 33, no. 6, 2009, pp. 539–567.

Stanton, B., J. Guo, L. Cottrell, J. Galbraith, X. Li, C. Gibson, et al. "The Complex Business of Adapting Effective Interventions to New Populations: An Urban to Rural Transfer." *Journal of Adolescent Health*, vol. 37, no. 2, 2005, pp. 163e17–163e26.

Stanton, B., C. Harris, L. Cottrell, X. Li, C. Gibson, J. Guo, et al. "Trial of an Urban Adolescent Sexual Risk-Reduction Intervention for Rural Youth: A Promising But Imperfect Fit." *Journal of Adolescent Health*, vol. 38, no. 1, 2006, pp. 55e25–55e36.

Suellentrop, K., and G. Hunter. "American Indian/Alaska Native Youth and Teen Pregnancy Prevention." Washington, DC: National Campaign to Prevent Teen and Unplanned Pregnancy, 2009.

Tortolero, S. R., C. M. Markham, M. Fleschler Peskin, R. Shegog, R. C. Addy, S. L. Escobar-Chavez, et al. "It's Your Game: Keep It Real: Delaying Sexual Behavior with an Effective Middle School Program." *Journal of Adolescent Health*, vol. 46, no. 2, 2009, pp. 1–11.

Underhill, K., P. Montgomery, and D. Operario. "Sexual Abstinence Only Programmes to Prevent HIV Infection in High Income Countries: Systematic Review." *BMJ*, vol. 335, Epub, 2007a, p. 248.

Underhill, K., D. Operario, and P. Montgomery. "Systematic Review of Abstinence-Plus HIV Prevention Programs in High-Income Countries." *PLoS Med*, vol. 4, 2007b, e275.

United Nations. "2011 Demographic Yearbook." New York: United Nations, 2012.

Valentine, J.C., A. Biglan, R. F. Boruch, et al. "Replication in Prevention Science." *Prevention Science*, vol. 12, 2011, pp. 103–117.

Villarruel, A. M., J. B. Jemmott, and L. S. Jemmott. "A Randomized Controlled Trial Testing an HIV Prevention Intervention for Latino Youth." *Archives of Pediatrics & Adolescent Medicine*, vol. 160, no. 8, 2006, pp. 772–777.

Weinstock, H., S. Berman, and W. Cates Jr. "Sexually Transmitted Diseases Among American Youth: Incidence and Prevalence Estimates, 2000." *Perspectives on Sexual and Reproductive Health*, vol. 36, 2004, pp. 6–10.

Welsh, B. C., C. J. Sullivan, and D. L. Olds. "When Early Crime Prevention Goes to Scale: A New Look at the Evidence." *Prevention Science*, vol. 11, no. 2, 2010, pp. 115–125.

What Works Clearinghouse. *Procedures and Standards Handbook (Version 2.1)*. Available at: [<http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19>]. Accessed August 11, 2012.