Journal of Adolescent Health 54 (2014) 499–507

ELSEVIER

JOURNAL OF ADOLESCENT HEALTH

www.jahonline.org

Review article

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

Brian Goesling, Ph.D. [a,*], Silvie Colman, Ph.D. [a], Christopher Trenholm, Ph.D. [a], Mary Terzian, Ph.D. [b], and Kristin Moore, Ph.D. [b]

[a] *Mathematica Policy Research, Princeton, New Jersey*
[b] *Child Trends, Bethesda, Maryland*

Article history: Received June 24, 2013; Accepted December 2, 2013
Keywords: Evidence-based programs; HIV; Sexually transmitted infections; Systematic review; Teen pregnancy

A B S T R A C T

Purpose: This systematic review provides a comprehensive, updated assessment of programs with evidence of effectiveness in reducing teen pregnancy, sexually transmitted infections (STIs), or associated sexual risk behaviors.
Methods: The review was conducted in four steps. First, multiple literature search strategies were used to identify relevant studies released from 1989 through January 2011. Second, identified studies were screened against prespecified eligibility criteria. Third, studies were assessed by teams of two trained reviewers for the quality and execution of their research designs. Fourth, for studies that passed the 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 programs delivered in diverse settings. Most studies had mixed-gender and predominately African-American research samples (70% and 51%, respectively). Randomized controlled trials accounted for the large majority (87%) of included studies. Most studies (76%) included multiple follow-ups, with sample sizes ranging from 62 to 5,244. Analysis of the study impact findings identified 31 programs with evidence of effectiveness.
Conclusions: Research conducted since the late 1980s has identified more than two dozen teen pregnancy and STI prevention programs with evidence of effectiveness. Key strengths of this research are the large number of randomized controlled trials, the common use of multiple follow-up periods, and attention to a broad range of programs 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, researchers must overcome common limitations in study design, analysis, and reporting that have negatively affected prior research.

IMPLICATIONS AND CONTRIBUTION

Researchers, policymakers, and practitioners need reliable information on the effectiveness of individual teen pregnancy prevention programs to help identify programs to consider for broader dissemination. To help meet this need, this review article identifies 31 individual programs with evidence of effectiveness in reducing teen pregnancy, sexually transmitted infections, or associated sexual risk behaviors. It also documents the relative strengths and weaknesses of the existing evidence and identifies priorities for future research.

Financial Disclosure: This research was supported by a U.S. Department of Health and Human Services contract with Mathematica Policy Research (HHSP23320095642WC); however, the views expressed here do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. government.
* Address correspondence to: Brian Goesling, Ph.D., Mathematica Policy Research, P.O. Box 2393, Princeton, NJ 08543-2393.
E-mail address: bgoesling@mathematica-mpr.com (B. Goesling).

High rates of teen pregnancy, sexually transmitted infections (STIs), and associated sexual risk behaviors remain a troubling issue in the United States. Nationwide, 24% of high school students report having had four or more partners by graduation, and nearly 40% of sexually active students had not used a condom during their last sexual intercourse [1]. These behaviors increase the risks of pregnancy and STIs, including HIV. Preliminary national data for 2012 indicate there were approximately 29.4

1054-139X/$ – see front matter
http://dx.doi.org/10.1016/j.jadohealth.2013.12.004

births per 1,000 females 15 to 19 years of age [2], a rate higher than in most other industrialized countries [3]. In addition, estimates suggest that adolescents and young adults account for half of all new STI cases in the United States every year [4].

Several prior reviews have summarized the effectiveness of programs intended to reduce these risks. Most of these reviews have sought to synthesize evidence across broad categories of programs—for example, abstinence education programs [5,6]; comprehensive sex education programs [5,6]; school-based prevention programs [7,8]; positive youth development programs [9]; or media-based approaches [10]. Within each of these broad categories, however, there exists a potentially wide range of individual programs, each with a slightly different focus and approach. Fewer studies have sought to identify and assess evidence separately for each individual program [11–13].

Assessing the evidence for individual programs is important for two reasons. First, two programs may have different effects even if they follow the same general approach. For example, a recent systematic review of comprehensive risk-reduction programs for adolescents found effects for individual programs ranging from favorable to adverse [5]. Averaging effects across individual programs within broader categories can mask such variation. Second, policymakers and practitioners need practical guidance in identifying individual programs to consider for broader dissemination. Especially given the potential for variation in program effects, research must provide guidance beyond the selection of broad categories of programs (such as abstinence education, comprehensive sex education, or positive youth development). The choice of program within each category may also matter.

The present study contributes to the field in several ways. First, we provide a comprehensive, updated assessment of individual programs with evidence of effectiveness in reducing teen pregnancy, STIs, or associated sexual risk behaviors. We identify and assess evidence separately for each individual program, without respect to the program's specific content or approach. We base this assessment on a systematic review of the literature, covering research published or released from 1989 through early 2011. Second, to provide context for this assessment, we also examine the relative strengths and weakness of the evidence. This context is important both to note potential limitations of our assessment and to identify priorities for future research.

Methods

The review was conducted following a prespecified protocol. We first developed the protocol in fall 2009 to identify and assess studies released from 1989 through January 2010. We later updated the protocol in fall 2010 to identify and assess newer studies released from January 2010 through January 2011. The findings presented in this study thus cover research released over a 22-year period from 1989 through January 2011. The review was sponsored by the U.S. Department of Health and Human Services (HHS) and conducted by researchers from Mathematica Policy Research and Child Trends. The review protocol is available for download on the HHS Web site [14].

Study identification

We identified studies in five ways: (1) scanning the reference lists of prior systematic reviews and research syntheses [11–13,15–17]; (2) searching the Web sites of relevant federal agencies and research or policy organizations; (3) issuing two public calls for studies to identify new or unpublished research; (4) having a research librarian conduct a keyword search of electronic citation databases; and (5) hand searching 10 relevant research journals and the conference proceedings of five professional associations. The entire search covered both published and unpublished studies. Focusing only on published studies can lead to bias in systematic reviews [18], because published studies tend to over-represent favorable and statistically significant findings relative to null or negative findings. Additional details on the search strategy are available in the review protocol [14].

Study screening

The study screening process had two steps. First, teams of two researchers screened the titles and abstracts of all studies identified through the literature search. Studies that did not meet the prespecified eligibility criteria (listed below) were excluded from the review. Second, for studies that passed the first stage of screening, the review team obtained full text of the identified reports and journal articles to conduct a second-stage screening. Findings from a single study presented in multiple reports or journal articles were linked and assessed together [e.g., 19–23]. In reports or journal articles presenting findings from multiple studies or a multiarmed trial, each study or separable trial arm was assessed separately [e.g., 24].

To be eligible for the review, a study had to meet four inclusion criteria. First, the study had to examine the impacts of an intervention using quantitative data and statistical analysis and hypothesis testing. Both randomized controlled trials and quasiexperiments were eligible. Second, 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). Third, the study sample had to consist of U.S. youth age 19 years or younger at the time of sample enrollment. Fourth, the range of eligible programs covered those intending to reduce rates of teen pregnancy, STIs, or associated sexual risk behaviors through any combination of educational, skill-building, and/or psychosocial intervention. We included both programs offering services one-on-one to individuals and those serving groups. Examples included classroom-based health curricula, individualized programs delivered by health professionals in clinics or other settings, community-based or after-school programs, and specialized programs for youth in the juvenile justice or child welfare systems.

Studies were excluded from the review if they were conducted outside the United States; if they measured program impacts only on composite scales of sexual risk behavior or measures without established validity (e.g., reports from males of their female partners' use of birth control pills); or if they assessed the impacts of the following types of interventions: early childhood education programs, home visiting programs, high school dropout prevention programs, or broad state- or federal-level policy changes.

Study quality assessment

All studies that met the review eligibility criteria were assessed by teams of two trained reviewers for the quality and execution of their research designs. The reviewers made their assessments using a modified version of the rating tool used by

the U.S. Department of Education's What Works Clearinghouse [25]. Differences of opinion were resolved through consensus.

As part of the rating tool used for the assessment, the reviewers assigned each study a final 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.

The moderate quality rating was considered for (1) quasiexperimental 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 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 typically low for this age group. As required for the highest study rating, to meet the review criteria for a moderate 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 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, because 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 [14].

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 the following information for each program impact estimate: the name and description of the outcome measure, length of follow-up, analytic sample used to estimate the program impact, 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. We extracted this impact information only for eligible outcome measures as defined in the review protocol.

Our review protocol also called for collecting effect size information to assess the magnitude of each reported impact estimate. However, we found that effect size information was often missing from the included studies, and that the information when reported was not directly comparable across studies due to differences in the study design, analysis methods, or metric used to measure the effect (odds ratio, relative risk, standardized mean difference, and so on). In part for these reasons, we did not ultimately collect effect size information for all included studies or require this information for the subsequent analysis.

Analysis

The analysis had two main steps. First, we tabulated all studies included in the data extraction by program type and other key features of the evaluation setting, study sample, and research design. We used these tabulations to identify the relative strengths and weaknesses of the evidence. Second, we then used the program impact findings to identify programs with evidence of effectiveness, defined as having a statistically significant positive impact (and no adverse effects) on at least one of the following outcomes: sexual activity; contraceptive use or consistency of use; STIs; or pregnancy or birth. To reduce the possibility of detecting chance findings due to multiple-hypothesis testing, we limited this assessment to program impacts estimated for either the full study sample or a subgroup defined by gender or baseline sexual experience.

In identifying programs with evidence of effectiveness, we did not employ "vote counting," meta-analysis, or any other method of synthesizing impact findings across multiple studies. As reported in the results section below, the large majority of teen pregnancy prevention programs have been evaluated only once. Our criteria for program effectiveness thus required findings from only a single impact study, and we did not have need for methods of synthesizing findings across multiple studies.

Throughout the analysis, the team did not consider evidence for 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 [26]. To minimize the risk of bias, we excluded such estimates from our analysis.

Results

We identified more than 1,900 citations through our literature search (Figure 1). From this initial citation list, we excluded 1,438 (73%) citations after screening on titles and abstracts. We obtained full text reports and journal articles for the remaining 541 citations, and from these citations, we identified 452 unique studies. We excluded an additional 252 studies after reading the full text, and 112 studies were excluded 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. A complete list of studies reviewed is available online [14].

The studies tested a mix of programs delivered in diverse settings (Table 1). Nearly half the included studies (47%) 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%), clinic-based programs offering individualized services (11%), youth development programs (11%), or programs for specialized populations such as pregnant or parenting teens or youth in the foster care or juvenile justice systems (11%). Most programs were delivered in after-school or community-based organizations (38%) or in school during the regular school day (29%). Among the in-school programs, more were delivered in middle schools than in high schools or elementary schools.

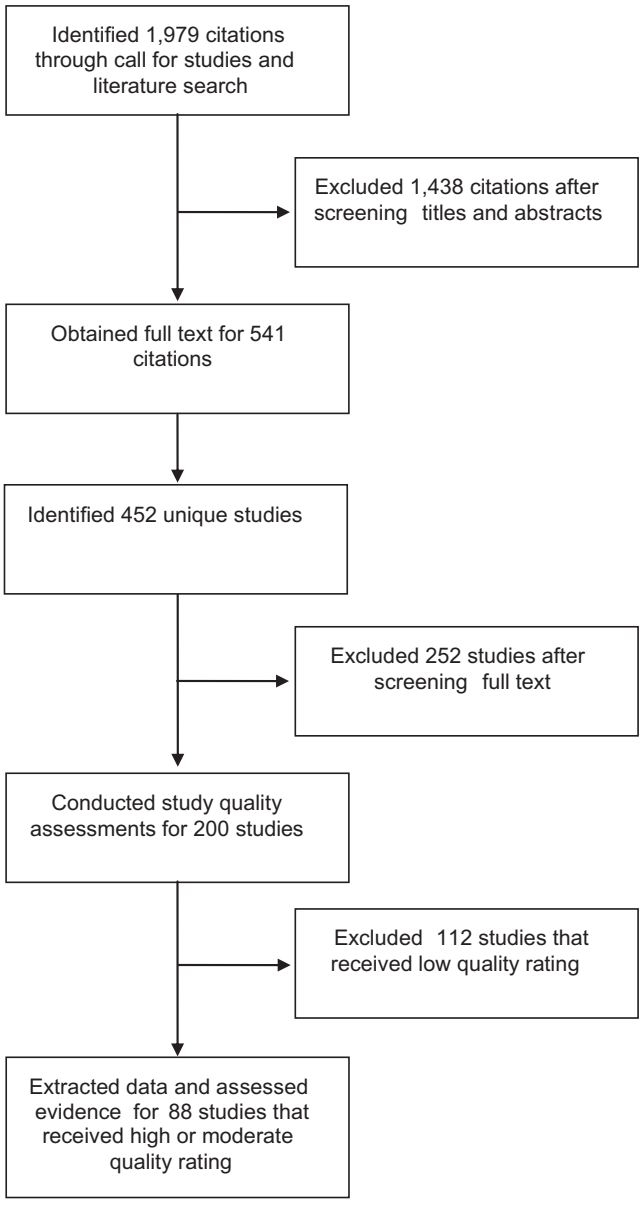

Figure 1. Flow of citations and studies through the review.

Relatively few studies were set in community health clinics (16%) or specialized settings such as juvenile justice facilities (11%).

Most studies (70%) included mixed-sex samples, and slightly more than half featured predominately African-American samples (51%). 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 13 years and under (44%) and youth ages 14 to 17 years (44%). Fewer studies focused on older youth ages 18 or 19 years (11%).

Randomized controlled trials accounted for a large majority of included studies (Table 2). The most common method was random assignment of individual youth (44%), but nearly an equal proportion (41%) assigned youth in clusters such as schools. Few studies used quasiexperimental designs (13%).

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[a]	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[b]	10 (11)
Average age group	
13 years or younger	39 (44)
14 to 17 years	39 (44)
18 or 19 years	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)

[a] 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.
[b] Comprises juvenile justice facilities, residential facilities for substance dependent youth, and other specialized settings.

Sample sizes ranged from a low of 62 to a high of 5,244, with a median sample size of 447.

Nearly half (47%) of the studies 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%) and contraceptive use and/or consistency (80%). Fewer studies examined impacts on STIs (23%) or pregnancies or births (28%).

From the 88 studies included in the data extraction, we identified 31 programs with evidence of effectiveness in reducing teen pregnancy, STIs, or associated sexual risk behaviors (Figure 2). Among the other programs assessed, 34 demonstrated no evidence of a statistically significant favorable impact on an eligible outcome measure for either the full sample or a key subgroup. Another 13 programs had evidence of favorable impacts only for an endogenous subgroup defined by sexual activity at follow-up. The number of programs assessed is only slightly lower than the number of included studies (78 programs vs. 88 included studies) because the large majority of programs were evaluated only once.

For the 31 programs meeting the review criteria for evidence of effectiveness, most of the favorable impact findings focused on measures of sexual activity and contraceptive use (Table 3). Of the 31 programs, 22 had impacts on a measure of sexual activity, 14 had impacts on a measure of contraceptive use or consistency, five

Table 2
Design characteristics of included studies (n = 88)

Characteristic	Number of studies (percentage)
Study design	
Randomized controlled trial	77 (87)
Cluster	36 (41)
Individual	39 (44)
Mixed[a]	2 (2)
Quasiexperimental 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[b]	
Shortest	0 months
Median	0 months
Longest	72 months
Length of last follow-up[b]	
Shortest	0 months
Median	12 months
Longest	180 months
Outcome measures[c]	
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)

[a] Some participants were randomly assigned in clusters and others as individuals.
[b] Measured as months since the end of the intervention.
[c] Percentages do not sum to 100 because some studies measure more than one outcome.

had impacts on STIs, and five had impacts on pregnancies or births. In addition, 20 programs 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 had impacts on all four categories of outcome measures. None of the programs have any evidence of adverse effects, a requirement of the review criteria.

Only one program had evidence of impacts replicated across multiple studies [31–33]. The other programs had evidence from a single impact study. In most cases, this impact study was a randomized controlled trial that met the review criteria for a high study-quality rating. The remaining programs had evidence from either a quasiexperimental study (four programs) or a randomized trial that met the review criteria for a moderate (but not high) study-quality rating (six programs).

Discussion

This systematic review provides a comprehensive, updated assessment of programs with evidence of effectiveness in reducing teen pregnancy, STIs, or associated sexual risk behaviors. To conduct this assessment, we identified and assessed some 200 program impact studies released from 1989 through January 2011. Of the studies assessed, 88 met the review criteria for study design and execution. Analysis of the study impact findings identified 31 programs with evidence of effectiveness. To provide context for these findings and identify the relative strengths and weaknesses of the evidence, we also examined the study design quality and other characteristics of all 88 studies included in the analysis.

Prominence of randomized controlled trials

A main strength of the evidence is the large number of randomized controlled trials. In some areas of program evaluation and policy research, randomized controlled trials are either not feasible or considered an unrealistic standard. In this review, however, a large majority of included studies (87%) used randomized designs. These designs have been used successfully with all types of programs and in diverse settings, ranging from schools [29,38,41] to juvenile justice facilities [55]. 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 evidence 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.

Future research should more carefully consider the best timing for follow-up surveys. Most of today's teen pregnancy and STI prevention programs are built on logic models predicting both shorter-term impacts on sexual activity or contraceptive use and longer-term impacts on pregnancies or STIs [11]. However, our review findings suggest that researchers do not always design program impact studies in ways that allow for testing program effectiveness across this full range of outcomes. For example, among the 31 programs meeting the review criteria for evidence of effectiveness, we found that relatively few were tested for impacts on longer-term outcomes such as pregnancies and STIs. To allow for the testing of program impacts on a broader range of outcomes, researchers should consider the trade-offs in alternative follow-up schedules. For example, instead of administering three relatively short-term surveys (e.g., at immediate post-test and at 6 and 12 months postintervention), researchers could instead administer two surveys, one shorter-term (e.g., 6 months postprogram) and one long-term (e.g., 18–24 months postprogram). For many programs, the latter schedule may allow for testing impacts on a broader range of outcomes at similar data collection cost.

Diversity of programs and settings

The evidence is also strong in its focus on a broad range of programs delivered in diverse settings. 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.

Table 3
Programs with evidence of effectiveness (n = 31)

Program	Sexual activity	Contraceptive use	Sexually transmitted infections	Pregnancy or birth
Aban Aya Youth Project [27]	+	na	na	na
Adult Identity Mentoring (Project AIM)[a] [28]	+	na	na	na
All4You! [29]	+	+	na	na
Assisting in Rehabilitating Kids (ARK)[a] [30]	+	+	na	na
Be Proud! Be Responsible![a] [31–33]	+	+	na	na
Be Proud! Be Responsible! Be Protective![a] [34]	+	o	na	na
Becoming a Responsible Teen (BART)[a] [35]	+	+	na	na
Children's Aid Society (CAS)—Carrera Program[a] [36]	+	na	na	+
¡Cuídate![a] [37]	+	+	na	na
Draw the Line/Respect the Line[a] [38]	+	na	na	na
FOCUS[a] [39]	+	o	na	na
Heritage Keepers Abstinence Education [40]	+	na	na	na
Horizons[a] [51]	na	+	+	na
It's Your Game: Keep it Real [41]	+	na	na	na
Making a Difference![a] [24]	+	o	na	na
Making Proud Choices![a] [24]	o	+	na	na
Project TALC[a] [56,57]	o	na	na	+
PHAT! Abstinence Only Intervention[a] [42]	+	o	na	na
PHAT! Comprehensive Abstinence and Safer Sex Intervention[a] [42]	+	o	na	na
Reducing the Risk [52]	o	+	na	o
Rikers Health Advocacy Program (RHAP) [53]	o	+	na	na
Raising Healthy Children [43–45]	+	o	+	+
Respeto/Proteger[a] [54]	na	+	na	na
Safer Choices[a] [22]	o	+	na	na
Safer Sex [46]	+	o	na	na
Sexual Health and Adolescent Risk Prevention (SHARP) [55]	na	+	na	na
SiHLE[a] [47]	+	+	+	+
Sisters Saving Sisters[a] [48]	+	+	+	na
Teen Health Project [49]	+	na	na	na
Teen Outreach Program[a] [58]	o	na	na	+
What Could You Do?[a] [50]	+	o	+	na

+ = statistically significant program impact; o = no statistically significant program impact; na = not available (either not measured or did not meet review criteria).
[a] Denotes programs supported by a randomized controlled trial that met the review criteria for a high rating.

This diversity is important for two reasons. For one, there is no single recipe for success in improving adolescent sexual health outcomes. Prior systematic reviews have shown that even within broad categories of similar programs (e.g., among all clinic-based or youth development programs), there is often significant variation in program impacts across individual programs [5,8,59,60]. Some individual programs have demonstrated evidence of success whereas others have not. Even in cases in which two individual programs have tried following a very similar approach, the impacts on youth outcomes have often differed (e.g., 36,61). Until there is more rigorous evidence about why some individual programs 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 may 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 [62,63]. Because no one size fits all, it is important to have a variety programs available for implementation, and this, in turn, requires a research literature that is equally broad in focus.

Need for increased diversity of target populations

The current evidence is relatively less diverse with respect to target population. Among the 88 studies included in the review, about half (51%) featured predominately African-American research samples and over three quarters (88%) focused on youth age 17 years or under. By contrast, comparatively few studies featured predominately Latino (19%) or white (28%) samples or focused on older youth ages 18 or 19 years (12%).

Diversity of target population is important to ensure the availability of effective programs for all types of youth. For example, although Latino youth currently have the highest teen birth rate of all major racial/ethnic groups in the United States [2], our review findings show that Latinos are comparatively under-represented in the evidence, and that only two programs designed specifically for use with Latino youth have demonstrated evidence of program effectiveness [37,54]. Similarly, there is currently limited evidence on effective programs for relatively small but high-risk groups such as pregnant or parenting teens [34,54], youth living in foster care [64], and American Indian and Alaska Natives, despite research showing that all of these groups are at above-average risk of teen pregnancy [65–67].

Need for studies of replication and scale-up

The biggest gap in the evidence is a lack of replication studies. Researchers increasingly recognize replication as a key step in the process of identifying effective interventions [68]. However, among the 31 programs we found to have evidence of effectiveness, only one program has shown favorable impacts across multiple studies [31–33]. Instead, most of the current evidence base consists of single small-scale "efficacy trials" [69] conducted

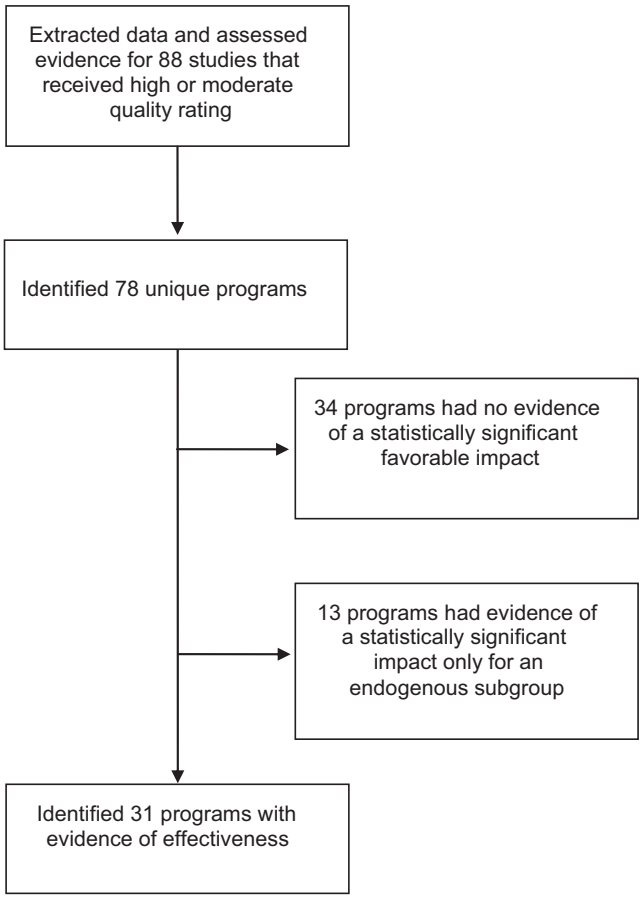

Figure 2. Identifying individual programs with evidence of effectiveness.

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, researchers must supplement initial efficacy studies with 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% when programs are implemented on a very large scale [70]. For the literature on teen pregnancy and STI prevention programs, these findings give warning that the existing evidence in support of some programs may weaken as the research literature expands.

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 quasiexperimental 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 [11].

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 [71,72]. 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.

Limitations

Possible limitations of the review include publication bias in the included studies, the use of author-reported statistical significance levels to identify programs with evidence of effectiveness, and missing information on program effect sizes. We addressed the issue of publication bias in part by including both published and unpublished studies in our literature search. However, researchers may have additional relevant findings for these programs that are currently unavailable in any public, written report. The reliance on author-reported statistical significance levels is a limitation because significance levels do not convey the magnitude or practical significance of the observed effects. In addition, some practically meaningful effects may be reported as nonsignificant if the study sample is small. These trade-offs were necessary, however, given the infrequent and inconsistent reporting of effect sizes or other measures of program effectiveness. Future updates to these review findings should seek to incorporate program effect size information if or when this information is more consistently reported in the research literature.

References

[1] Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance—United States, 2011. MMWR 2012;61(SS-4):1–162.
[2] Hamilton BE, Martin JA, Ventura SJ. Births: Preliminary data for 2012. National Vital Statistics Reports, vol. 62, no. 3. Hyattsville, MD: National Center for Health Statistics; September 2013.
[3] United Nations. 2011 demographic yearbook. New York: United Nations; 2012.
[4] Centers for Disease Control and Prevention. Sexually transmitted disease surveillance 2011. Atlanta, GA: U.S. Department of Health and Human Services; 2012.
[5] Chin HB, Sipe TA, Elder R, 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. Am J Prev Med 2012;42:272–94.
[6] Kohler PK, Manhart LE, Lafferty WE. Abstinence-only and comprehensive sex education and the initiation of sexual activity and teen pregnancy. J Adolesc Health 2008;42:341–51.
[7] Bennett SE, Assefi NP. School-based teenage pregnancy prevention programs: A systematic review of randomized controlled trials. Journal of Adolescent Health 2005;36:72–81.
[8] Silva M. The effectiveness of school-based sex education programs in the promotion of abstinent behavior: A meta-analysis. Health Educ Res 2002;17:471–81.

[9] Gavin LE, Catalano RF, David Ferdon C, et al. A review of positive youth development programs that promote adolescent sexual and reproductive health. Journal of Adolescent Health 2010;46:S75–91.
[10] Guse K, Levine D, Martins S, et al. Interventions using new digital media to improve adolescent sexual health: A systematic review. Journal of Adolescent Health 2012;51:535–43.
[11] 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.
[12] Advocates for Youth. Science and success. 2nd ed. Washington, DC: Advocates for Youth; 2008.
[13] Kim CC, Rector R. Abstinence education: Assessing the evidence. Washington, DC: The Heritage Foundation; 2008.
[14] Mathematica Policy Research. Identifying programs that impact teen pregnancy, sexually transmitted infections, and associated sexual risk behaviors: Review protocol. Available at: http://www.hhs.gov/ash/oah/oah-initiatives/teen_pregnancy/db/.
[15] Scher L, Maynard RA, Stagner M. Interventions intended to reduce pregnancy-related outcomes among adolescents. Campbell Systematic Reviews 2006;12.
[16] Ball V, Moore KA. What works for adolescent reproductive health: Lessons from experimental evaluations of programs and interventions. Washington, DC: Child Trends; 2008.
[17] Oringanje C, Meremikwu MM, Eko H, et al. Interventions for preventing unintended pregnancies among adolescents. Cochrane Database Syst Rev 2009.
[18] Institute of Medicine (IOM). Finding what works in health care: Standards for systematic reviews. Washington, DC: The National Academies Press; 2011.
[19] Basen-Engquist K, Coyle K, Parcel G, et al. Schoolwide effects of a multicomponent HIV, STD, and pregnancy prevention program for high school students. Health Education & Behavior 2001;28:166–85.
[20] Coyle K, Basen-Engquist K, Kirby D, et al. Safer choices: Reducing teen pregnancy, HIV, and STDs. Public Health Rep 2001;116:82–93.
[21] Coyle K, Basen-Engquist K, Kirby D, et al. Short-term impact of safer choices: A multicomponent, school-based HIV, other STD, and pregnancy prevention program. J Sch Health 1999;69:181–8.
[22] Kirby DB, Baumler E, Coyle KK. The impact of "safer choices" on condom use and contraceptive use among sexually experienced students at baseline. [unpublished manuscript]; 2011.
[23] Kirby DB, Baumler E, Coyle KK, et al. The "safer choices" intervention: Its impact on the sexual behaviors of different subgroups of high school students. Journal of Adolescent Health 2004;35:442–52.
[24] Jemmott JB, Jemmott LS, Fong GT. Abstinence and safer sex HIV risk-reduction interventions for African-American adolescents: A randomized controlled trial. JAMA: Journal of the American Medical Association 1998;279:1529–36.
[25] What Works Clearinghouse. Procedures and standards handbook (Version 2.1). Available at: http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19.
[26] 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; 2012.
[27] Flay BR, Graumlich S, Segawa E, et al. Effects of 2 prevention programs on high-risk behaviors among African-American youth: A randomized trial. Archives of Pediatrics & Adolescent Medicine 2004;158:377–84.
[28] Clark LF, Miller KS, Nagy SS, et al. Adult identity mentoring: Reducing sexual risk for African-American seventh grade students. Journal of Adolescent Health 2005;37:337e1–337e10.
[29] Coyle KK, Kirby DB, Robin LE, et al. All4You! A randomized trial of an HIV, other STDs, and pregnancy prevention intervention for alternative school students. AIDS Educ Prev 2006;18:187–203.
[30] St Lawrence JS, Crosby RA, Brasfield TL, O'Bannon III RE. Reducing STD and HIV risk behavior of substance-dependent adolescents: A randomized controlled trial. J Consult Clin Psychol 2002;70:1010–21.
[31] Jemmott III JB. Reductions in HIV risk-associated sexual behaviors among black male adolescents: Effects of an AIDS prevention intervention. Am J Public Health 1992;82:372–7.
[32] Jemmott JB, Jemmott LS, Fong GT, McCaffree K. Reducing HIV risk-associated sexual behavior among African-American adolescents: Testing the generality of intervention effects. Am J Community Psychol 1999;27:161–87.
[33] Jemmott III JB, Jemmott LS, Fong GT, Morales KH. 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 2010;100:720–6.
[34] Koniak-Griffin D, Lesser J, Nyamathi A, et al. Project CHARM: An HIV prevention program for adolescent mothers. Fam Community Health 2003;26:94–107.
[35] St Lawrence JS, Brasfield TL, Jefferson KW, et al. Cognitive-behavioral intervention to reduce African-American adolescents' risk for HIV infection. Journal of Consulting and Clinical Psychology 1995;63:221–37.
[36] Philliber S, Kaye JW, Herrling S, West E. 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 2002;34:244–51.
[37] Villarruel AM, Jemmott JB, Jemmott LS. A randomized controlled trial testing an HIV prevention intervention for Latino youth. Archives of Pediatrics & Adolescent Medicine 2006;160:772–7.
[38] Coyle KK, Kirby DB, Marin BV, et al. Draw the line/respect the line: A randomized trial of a middle school intervention to reduce sexual risk behaviors. American Journal of Public Health 2004;94:843–51.
[39] Boyer CB, Shafer MA, Shaffer RA, et al. Evaluation of a cognitive-behavioral, group, randomized controlled intervention trial to prevent sexually transmitted infections and unintended pregnancies in young women. Prev Med 2005;40:420–31.
[40] Weed SE, Birch PJ, Ericksen IH, Olsen JA. Testing a predictive model of youth sexual intercourse initiation. [unpublished manuscript]; 2011.
[41] Tortolero SR, Markham CM, Fleschler Peskin M, et al. It's Your Game: Keep It Real: Delaying sexual behavior with an effective middle school program. Journal of Adolescent Health 2009;46:1–11.
[42] Jemmott JB, Jemmott LS, Fong GT. Efficacy of a theory-based abstinence-only intervention over 24 months: A randomized controlled trial with young adolescents. Archives of Pediatrics & Adolescent Medicine 2010;164:152–9.
[43] Hawkins JD, Catalano RF, Kosterman R, et al. Preventing adolescent health-risk behaviors by strengthening protection during childhood. Archives of Pediatrics & Adolescent Medicine 1999;153:226–34.
[44] Hawkins JD, Kosterman R, Catalano RF, et al. Effects of social development intervention in childhood 15 years later. Archives of Pediatrics & Adolescent Medicine 2008;162:1133–41.
[45] Lonczak HS, Abbott RD, Hawkins JD, et al. 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 2002;156:438–47.
[46] Shrier LA, Ancheta R, Goodman E, et al. Randomized controlled trial of a safer sex intervention for high-risk adolescent girls. Archives of Pediatrics & Adolescent Medicine 2001;155:73–9.
[47] DiClemente RJ, Wingood GM, Harrington KF, et al. Efficacy of an HIV prevention intervention for African-American adolescent girls: A randomized controlled trial. JAMA: Journal of the American Medical Association 2004;292:171–9.
[48] Jemmott JB, Jemmott LS, Braverman PK, Fong GT. HIV/STD risk reduction interventions for African-American and Latino adolescent girls at an adolescent medicine clinic: A randomized controlled trial. Archives of Pediatrics & Adolescent Medicine 2005;159:440–9.
[49] Sikkema KJ, Anderson ES, Kelly JA, et al. Outcomes of a randomized, controlled community-level HIV prevention intervention for adolescents in low-income housing developments. AIDS 2005;19:1509–16.
[50] Downs JS, Murray PJ, Bruine de Bruin W, et al. Interactive video behavioral intervention to reduce adolescent females' STD risk: A randomized controlled trial. Social Science & Medicine 2004;59:1561–72.
[51] DiClemente RJ, Wingood GM, Rose ES, et al. Efficacy of sexually transmitted disease/human immunodeficiency virus sexual risk-reduction intervention for African-American adolescent females seeking sexual health services: A randomized controlled trial. Archives of Pediatrics & Adolescent Medicine 2009;163:1112–21.
[52] Kirby D, Barth RP, Leland N, Fetro JV. Reducing the risk: Impact of a new curriculum on sexual risk-taking. Fam Plann Perspect 1991;23:253–63.
[53] Magura S, Kang SY, Shapiro JL. Outcomes of intensive AIDS education for male adolescent drug users in jail. Journal of Adolescent Health 1994;15:457–63.
[54] Lesser J, Koniak-Griffin D, Huang R, et al. Parental protectiveness and unprotected sexual activity among Latino adolescent mothers and fathers. AIDS Education & Prevention 2009;21:88–102.
[55] Bryan AD, Schmiege SJ, Broaddus MR. HIV risk reduction among detained adolescents: A randomized, controlled trial. Pediatrics 2009;124:e1180–8.
[56] Rotheram-Borus MJ, Lee M, Leonard N, et al. Four-year behavioral outcomes of an intervention for parents living with HIV and their adolescent children. AIDS 2003;17:1217–25.
[57] Rotheram-Borus MJ, Lee M, Lin YY, Lester P. Six-year intervention outcomes for adolescent children of parents with the human immunodeficiency virus. Archives of Pediatric & Adolescent Medicine 2004;158:742–8.
[58] Allen JP, Philliber S, Herrling S, Kuperminc GP. Preventing teen pregnancy and academic failure: Experimental evaluation of a developmentally based approach. Child Development 1997;68:729–42.
[59] Underhill K, Montgomery P, Operario D. Sexual abstinence only programmes to prevent HIV infection in high income countries: Systematic review. BMJ 2007a;335:248.

[60] Underhill K, Operario D, Montgomery P. Systematic review of abstinence-plus HIV prevention programs in high-income countries. PLoS Med 2007b; 4:e275.
[61] Kirby D, Rhodes T, Campe S. 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.
[62] Stanton B, Guo J, Cottrell L, et al. The complex business of adapting effective interventions to new populations: An urban to rural transfer. Journal of Adolescent Health 2005;37:163e17–26.
[63] Stanton B, Harris C, Cottrell L, et al. Trial of an urban adolescent sexual risk-reduction intervention for rural youth: A promising but imperfect fit. Journal of Adolescent Health 2006;38:55e25–36.
[64] Kerr DCR, Leve LD, Chamberlain P. Pregnancy rates among juvenile justice girls in two randomized controlled trials of multidimensional treatment foster care. Journal of Counseling and Clinical Psychology 2009;77:588–93.
[65] Bilaver LA, Courtney ME. Foster care youth. Washington, DC: National Campaign to Prevent Teen Pregnancy; 2006.
[66] Klerman LV. Another chance: Preventing additional births to teen mothers. Washington, DC: National Campaign to Prevent Teen Pregnancy; 2004.
[67] Suellentrop K, Hunter G. American Indian/Alaska Native youth and teen pregnancy prevention. Washington, DC: National Campaign to Prevent Teen and Unplanned Pregnancy; 2009.
[68] Valentine JC, Biglan A, Boruch RF, et al. Replication in prevention science. Prev Sci 2011;12:103–17.
[69] Flay BR, Biglan A, Boruch RF, et al. Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. Prevention Science 2005;6: 151–75.
[70] Welsh BC, Sullivan CJ, Olds DL. When early crime prevention goes to scale: A new look at the evidence. Prevention Science 2010;11:115–25.
[71] Donner A, Klar N. Pitfalls of and controversies in cluster randomized trials. American Journal of Public Health 2004;94:416–22.
[72] Schochet P. An approach for addressing the multiple testing problem in social policy impact evaluations. Education Review 2009;33: 539–67.