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

ABSTRACT

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.

© 2014 Society for Adolescent Health and Medicine. All rights reserved.

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 (HHS023120095642WC); 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
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 at 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.
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%). 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). 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>
<thead>
<tr>
<th>Characteristic</th>
<th>Number of studies (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program type</td>
<td></td>
</tr>
<tr>
<td>Abstinence-based</td>
<td>17 (19)</td>
</tr>
<tr>
<td>Clinic-based</td>
<td>10 (11)</td>
</tr>
<tr>
<td>Sexuality education</td>
<td>41 (47)</td>
</tr>
<tr>
<td>Programs for special populations a</td>
<td>10 (11)</td>
</tr>
<tr>
<td>Youth development</td>
<td>10 (11)</td>
</tr>
<tr>
<td>Program length</td>
<td></td>
</tr>
<tr>
<td>Fewer than 10 sessions</td>
<td>57 (65)</td>
</tr>
<tr>
<td>10 to 20 sessions</td>
<td>13 (15)</td>
</tr>
<tr>
<td>More than 20 sessions</td>
<td>18 (20)</td>
</tr>
<tr>
<td>Evaluation setting</td>
<td></td>
</tr>
<tr>
<td>After-school/community-based</td>
<td>33 (38)</td>
</tr>
<tr>
<td>Health clinic</td>
<td>14 (16)</td>
</tr>
<tr>
<td>In-school</td>
<td>26 (29)</td>
</tr>
<tr>
<td>Elementary school</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Middle school</td>
<td>18 (20)</td>
</tr>
<tr>
<td>High school</td>
<td>6 (7)</td>
</tr>
<tr>
<td>Multiple settings</td>
<td>5 (6)</td>
</tr>
<tr>
<td>Specialized setting b</td>
<td>10 (11)</td>
</tr>
<tr>
<td>Average age group</td>
<td></td>
</tr>
<tr>
<td>13 years or younger</td>
<td>39 (44)</td>
</tr>
<tr>
<td>14 to 17 years</td>
<td>39 (44)</td>
</tr>
<tr>
<td>18 or 19 years</td>
<td>10 (11)</td>
</tr>
<tr>
<td>Majority racial/ethnic group</td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>45 (51)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Latino</td>
<td>17 (19)</td>
</tr>
<tr>
<td>White</td>
<td>25 (28)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Both sexes</td>
<td>62 (70)</td>
</tr>
<tr>
<td>Female only</td>
<td>19 (22)</td>
</tr>
<tr>
<td>Male only</td>
<td>7 (8)</td>
</tr>
</tbody>
</table>

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.
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 quasiresearch 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 [53]. 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.
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 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 organized 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...
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 [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


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.


Weed SE, Birch PJ, Erickson IH, Olsen JA. Testing a predictive model of youth sexual intercourse initiation. [unpublished manuscript]; 2011.


Underhill K, Montgomery P, Operario D. Sexual abstinence only programs to prevent HIV infection in high income countries: Systematic review. BMJ 2007a;335:248.