Cochrane Library

Cochrane Database of Systematic Reviews

Abstinence-only programs for HIV infection prevention in high-income countries (Review)

Underhill K, Operario D, Montgomery P

Underhill K, Operario D, Montgomery P.
Abstinence-only programs for HIV infection prevention in high-income countries.
Cochrane Database of Systematic Reviews 2007, Issue 4. Art. No.: CD005421.
DOI: 10.1002/14651858.CD005421.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Abstinence-only programs for HIV infection prevention in high-income countries

Kristen Underhill[2], Don Operario[1], Paul Montgomery[3]

[1]Department of Social Policy and Social Work, University of Oxford, Oxford, UK. [2]Yale Law School, New Haven, CT, USA. [3]The Centre for Evidence-Based Intervention, University of Oxford, Oxford, UK

Contact address: Don Operario, Department of Social Policy and Social Work, University of Oxford, Oxford, UK. don.operario@applied-social-studies.oxford.ac.uk.

Editorial group: Cochrane HIV/AIDS Group.
Publication status and date: Edited (no change to conclusions), published in Issue 1, 2009.
Review content assessed as up-to-date: 21 August 2007.

Citation: Underhill K, Operario D, Montgomery P. Abstinence-only programs for HIV infection prevention in high-income countries. *Cochrane Database of Systematic Reviews* 2007, Issue 4. Art. No.: CD005421. DOI: 10.1002/14651858.CD005421.pub2.

ABSTRACT

Background

Abstinence-only interventions promote sexual abstinence as the only means of preventing sexual acquisition of HIV; they do not promote safer-sex strategies (e.g., condom use). Although abstinence-only programs are widespread, there has been no internationally focused review of their effectiveness for HIV prevention in high-income countries.

Objectives

To assess the effects of abstinence-only programs for HIV prevention in high-income countries.

Search methods

We searched 30 electronic databases (e.g., CENTRAL, PubMed, EMBASE, AIDSLINE, PsycINFO) ending February 2007. Cross-referencing, handsearching, and contacting experts yielded additional citations through April 2007.

Selection criteria

We included randomized and quasi-randomized controlled trials evaluating abstinence-only interventions in high-income countries (defined by the World Bank). Interventions were any efforts to encourage sexual abstinence for HIV prevention; programs that also promoted safer-sex strategies were excluded. Results were biological and behavioral outcomes.

Data collection and analysis

Three reviewers independently appraised 20,070 records and 326 full-text papers for inclusion and methodological quality; 13 evaluations were included. Due to heterogeneity and data unavailability, we presented the results of individual studies instead of conducting a meta-analysis.

Main results

Studies involved 15,940 United States youth; participants were ethnically diverse. Seven programs were school-based, two were community-based, and one was delivered in family homes. Median final follow-up occurred 17 months after baseline.

Results showed no indications that abstinence-only programs can reduce HIV risk as indicated by self-reported biological and behavioral outcomes. Compared to various controls, the evaluated programs consistently did not affect incidence of unprotected vaginal sex, frequency of vaginal sex, number of partners, sexual initiation, or condom use.

One study found a significantly protective effect for incidence of recent vaginal sex (n=839), but this was limited to short-term follow-up, countered by measurement error, and offset by six studies with non-significant results (n=2615).

One study found significantly harmful effects for STI incidence (n=2711), pregnancy incidence (n=1548), and frequency of vaginal sex (n=338); these effects were also offset by studies with non-significant findings.

Methodological strengths included large samples, efforts to improve self-report, and analyses controlling for baseline values. Weaknesses included underutilization of relevant outcomes, underreporting of key data, self-report bias, and analyses neglecting attrition and clustered randomization.

Authors' conclusions

Evidence does not indicate that abstinence-only interventions effectively decrease or exacerbate HIV risk among participants in high-income countries; trials suggest that the programs are ineffective, but generalizability may be limited to US youth. Should funding continue, additional resources could support rigorous evaluations with behavioral or biological outcomes. More trials comparing abstinence-only and abstinence-plus interventions are needed.

PLAIN LANGUAGE SUMMARY

Abstinence-only programs for preventing HIV infection in high-income countries (as defined by the World Bank)

Abstinence-only programs are widespread and well-funded, particularly in the United States and countries supported by the US President's Emergency Plan for AIDS Relief. On the premise that sexual abstinence is the best and only way to prevent HIV, abstinence-only interventions aim to prevent, stop, or decrease sexual activity. These programs differ from abstinence-plus designs: abstinence-plus programs promote safer-sex strategies (e.g., condom use) along with sexual abstinence, but abstinence-only programs do not, and instead often highlight the limitations of condom use. An up-to-date review suggests that abstinence-only programs do not affect HIV risk in low-income countries; this review examined the evidence in high-income countries.

This review included thirteen randomized controlled trials comparing abstinence-only programs to various control groups (e.g., "usual care," no intervention). Although we conducted an extensive international search for trials, all included studies enrolled youth in the US (total baseline enrollment=15,940 participants). Programs were conducted in schools, community centers, and family homes; all were delivered in family units or groups of young people. We could not conduct a meta-analysis because of missing data and variation in program designs. However, findings from the individual trials were remarkably consistent.

Overall, the trials did not indicate that abstinence-only programs can reduce HIV risk as indicated by behavioral outcomes (e.g., unprotected vaginal sex) or biological outcomes (e.g., sexually transmitted infection). Instead, the programs consistently had no effect on participants' incidence of unprotected vaginal sex, frequency of vaginal sex, number of sex partners, sexual initiation, or condom use.

One trial favored an abstinence-only program over usual care for incidence of vaginal sex (n=839), but this was limited to two-month follow-up and was offset by measurement error and six other studies with non-significant effects (n=2615).

One evaluation found several significant adverse (harmful) program effects: abstinence-only program participants were more likely than usual-care controls to report sexually transmitted infections (n=2711), pregnancy (n=1548), and increased frequency of vaginal sex (n=338). These effects were offset by high attrition and other studies showing non-significant effects.

We concluded that abstinence-only programs do not appear to reduce or exacerbate HIV risk among participants in high-income countries, although this evidence might not apply beyond US youth. Trial limitations included underreporting of relevant outcomes, reliance on program participants to report their behaviors accurately, and methodological weaknesses in the trials.

BACKGROUND

Although over two decades have passed since the first AIDS diagnosis, an effective and accessible HIV vaccine remains a distant hope. More than 7,600 people died from AIDS-related causes each day in 2005, while an estimated 38.6 million people worldwide were living with HIV (UNAIDS 2006). Poverty and structural violence, insufficient prevention efforts, and rapid viral evolution contribute to the spread of this "modern plague" (Farmer 2001); approximately 4.1 million new infections occurred in 2005 alone (UNAIDS 2006). In the absence of a vaccine, HIV prevention programs demand continued attention.

The HIV pandemic is most devastating for middle- and low-income nations, but new infections continue mounting even in countries with many resources for prevention. The World Health Organization estimated in 2004 that 1.6 million people in high-income countries were living with HIV (UNAIDS 2004); by 2005, 2.0 million individuals in North America, Western Europe, and Central Europe alone were living with HIV, and 65,000 became newly infected in these three regions (UNAIDS 2006). The widespread availability of antiretroviral treatment (ART) allows many individuals in high-income countries to live longer and healthier lives than their counterparts in resource-poor countries, but ART is not a cure. Furthermore, even if ART could eliminate the virus, it would still be desirable to prevent the spread of HIV. Primary prevention efforts are still necessary in high-income countries, particularly among high-risk groups.

High-income economies are defined by the World Bank as those with a gross national income per capita of $10,726 or higher: Andorra, Antigua and Barbuda, Aruba, Australia, Austria, the Bahamas, Bahrain, Belgium, Bermuda, Brunei Darussalam, Canada, Cayman Islands, Channel Islands, Cyprus, Denmark, Faeroe Islands, Finland, France, French Polynesia, Germany, Greece, Greenland, Guam, Hong Kong (China), Iceland, Ireland, Isle of Man, Israel, Italy, Japan, Korea Rep., Kuwait, Liechtenstein, Luxembourg, Macao (China), Malta, Monaco, Netherlands, Netherlands Antilles, New Caledonia, New Zealand, Norway, Portugal, Puerto Rico, Qatar, San Marino, Saudi Arabia, Singapore, Slovenia, Spain, Sweden, Switzerland, United Arab Emirates, United Kingdom, United States, US Virgin Islands (World Bank 2007).

Sexual behavior outpaces drug injecting as the major cause of HIV infection in many high-income countries. Studies from the United States suggest that over 70% of HIV-positive men and women may remain sexually active, and a substantial percentage continue to engage in unprotected sex (Crepaz 2002). The rising prevalence of other sexually transmitted infections (STIs), such as gonorrhea and chlamydia, indicates increased risky sexual behavior in Australia, Western Europe, Japan, and the US. Although sex between men still accounts for most new infections in Australia, Canada, Denmark, Germany, Greece, New Zealand, and the US, evidence suggests that heterosexual sex is responsible for an increasing proportion of seroconversions (UNAIDS 2004). For example, according to UK figures, heterosexual sex accounted for 66% of all new infections in 2003 (AVERT 2005). Major risk groups vary by region (Rivers 2000), but evidence suggests that HIV is now disproportionately concentrated among youth, ethnic minorities, women, recent immigrants, and men who have sex with men (UNAIDS 2006, Berry 2005, AVERT 2005). In the current socio-epidemiological climate, primary prevention programs targeting sexual risk behavior are essential for all risk groups.

Description of the intervention

Various behavioral interventions have proven effective in reducing HIV-risk behaviors in high-income countries, but a number of prevention strategies still require rigorous evaluation. Abstinence-only programs are one such approach.

Abstinence-only programs primarily encourage participants to refrain from sexual activity. They are designed to teach the social, health-related, and psychological benefits of abstaining from sexual activity; at the same time, most programs note the potential harms of sexual activity outside marriage (Devaney 2002, Haskins 1997, Rector 2002, Thomas 2000, Young 1998). They often target family involvement and community norms, as well as individual behaviors, addressing multiple influences on knowledge, attitudes, and values (Devaney 2002, Moore 1997, Young 1998). Their theoretical underpinnings include social learning theory, the health-belief model, cognitive-behavioral theory, the theory of social inoculation, the culture of poverty perspective, and utility maximization perspectives (Devaney 2002, Moore 1997, Thomas 2000). The interventions encourage both primary abstinence (remaining a virgin) and secondary abstinence (returning to abstinence after sexual activity) (Thomas 2000). The primary targets of these programs are young people who may not yet have initiated sexual activity.

Unlike abstinence-plus programs, abstinence-only interventions emphasize abstinence as the exclusive means of HIV prevention. If abstinence-only programs discuss contraception or condom use, they generally highlight the limitations of these strategies (Besharov 1997, Brown 1997, HRW 2002). In contrast, abstinence-plus programs promote abstinence as the best HIV prevention option, but also actively promote condom use and other safer-sex strategies. This review is limited to abstinence-only interventions, as a counterpart to an upcoming review of abstinence-plus interventions in high-income countries.

Terminology complicates discussions of abstinence-only programs. Specific definitions of "abstinence" and "sex" differ, and these terms are not always defined clearly (Bailey 2002). In a sur-

vey of federally funded abstinence-based education programs in all 50 US states, only six program coordinators reported defining "sexual activity" for program purposes (Sonfield 2001). "Sex" and "abstinence" also have many definitions among clinicians and the general public (Haglund 2002, Horan 1998, Pitts 2001, Remez 2000, Sanders 1999). But despite this variation, abstinence-only programs share the goal of discouraging sexual activity rather than making it safer. This common interest permits a systematic review.

Why it is important to do this review

Existing evaluations of abstinence-only programs are often US-focused, mixed in results, and inconsistent in methodology. For example, one review of abstinence-only programs reported significant reductions in sexual activity (Rector 2002); however, it was limited to US adolescents and included studies not designed to assess effectiveness (e.g., survey-based longitudinal studies without control groups). In contrast, another review concluded that "no evaluated program with an exclusive abstinence message has been evaluated in such a way as to show a significantly positive impact on behavior" (Thomas 2000); however, this review focused on adolescent pregnancy prevention, analyzed only programs endorsed by the US federal government, and included non-controlled studies. An exceptionally rigorous review of pregnancy prevention programs has also included that there is little to no evidence for the effectiveness of abstinence-only interventions (Kirby 2001). Many reviews include other types of programs alongside abstinence-only interventions, making separation difficult (DiCenso 2002, Pedlow 2003, Robin 2004). Reviews may also be limited to adolescents and may focus only on pregnancy prevention instead of HIV risk reduction (Franklin 1997, Kirby 2001). Although further analyses of abstinence-based programs are planned in high-income countries, these appear to be limited to adolescent groups in the US.

Abstinence-only programs have received considerable political attention, particularly in the US, making objective assessment difficult. As a strategy to reduce teen pregnancy, abstinence-only programs received nearly $900 million in US funding between 1999 and 2005, of which $168 million in federal funds was allocated for 2005 (Connolly 2004). However, the issue is polarized. For example, Representative Henry A Waxman reviewed the 13 most popular abstinence-only curricula receiving federal funding; this 2004 report stated that over 80% contained "false, misleading, or distorted information about reproductive health" (Waxman 2004). Groups such as Focus on the Family and the Government Affairs for Family Research Council disagreed, suggesting that the report "attempt[ed] to discredit highly successful abstinence programs" (Phan 2004). The debate has intensified during the summer of 2007, which has seen several Congressional decisions regarding abstinence-only program funding. In this setting, a systematic and apolitical review of effects without geographical limitations is essential to determine the utility of abstinence-only interventions.

To date, there has been no systematic analysis of the effects of abstinence-only programs on HIV prevention among all residents of high-income countries. This review seeks to identify, synthesize, and evaluate the effects of abstinence-only interventions on HIV-risk behavior and HIV transmission among participants in high-income countries.

OBJECTIVES

To determine the effects of abstinence-only programs for preventing HIV infection among participants in high-income countries.

The interventions were any planned efforts intended to increase rates of abstinence as an exclusive means of HIV prevention. The participants were anyone in high-income countries. Included outcomes were behavioral and biological. We included randomized and quasi-randomized controlled trials comparing abstinence-only programs to any control groups. These included no intervention; HIV-unrelated programs comparable in time and format (i.e., "attention controls;" for example, these might include a program involving the same activities and the same number of sessions, but focused on abstinence from drugs, not sex); abstinence-plus HIV prevention programs; and HIV prevention programs that did not emphasize abstinence.

Acknowledgement: much of our methodology and many of our identifying criteria have been reproduced with permission from the Cochrane protocol by Sweat, et al, "Abstinence-only programs for preventing HIV in developing countries" (also see O'Reilly 2004 and O'Reilly 2006).

METHODS

Criteria for considering studies for this review

Types of studies

Following the Cochrane Handbook and the precedent of previous HIV-AIDS related reviews (Siegfried 2003, Grimwade 2003), we included only randomized and quasi-randomized controlled trials. Quasi-randomized controlled trials had randomization sequences that were unlikely to lead to consistent bias, such as allocation by date of birth, coin flip, or alternation of individuals.

Included trials evaluated the effects of abstinence-only programs designed to influence behavior change on at least one behavioral or biological outcome measure related to HIV transmission. This limitation was practical because many of the primary outcomes were self-reported, and random allocation is important to control for both known and unknown confounders. If a meta-analysis had

been feasible, we would have conducted a separate analysis on randomized controlled trials to determine the effects of methodological quality.

Types of participants

We included studies comprising participants in high-income countries, as defined by the World Bank. Participants did not need to be born in or hold citizenship in a high-income country, but they must have been present in a high-income country when the intervention took place. No exclusions were made by intervention setting (e.g., clinic, school, community center, faith-based organization) or primary risk group. No exclusions were made by gender, age, sexual orientation, language, occupation, or racial or ethnic group.

Since our focus was primary prevention, studies restricted to participants who were already HIV-positive were excluded.

Types of interventions

"Abstinence" may be clinically defined as refraining from vaginal, anal, and oral sex. Given inconsistent definitions, this review included studies encouraging abstinence from any one of these behaviors or a combination of these behaviors. We made no exclusions by type of organization delivering the programs.

Criteria for abstinence-only interventions included the following:

1. the intervention was a planned effort to encourage sexual abstinence or a return to sexual abstinence;
2. specific outcomes of interest were presented;
3. HIV prevention was a stated goal of the intervention;
4. the program did not promote condom or contraception use (i.e., it had an exclusive abstinence focus).

We made no exclusions by type of comparison group. These could have included the following examples:

1. no intervention;
2. attention control: interventions that were equal in format and time, but targeted HIV-unrelated behaviors; for example, a comparison group might have received an intervention with the same number of sessions and the same activities, but the control intervention could have focused on refusing gang membership instead of abstaining from sexual activity;
3. interventions that did not encourage abstinence as a primary outcome (e.g., condom promotion programs);
4. abstinence-plus programs;
5. comparisons between enhanced and non-enhanced versions of the same program;
6. usual care as defined by the trialist.

Trials of programs that focused exclusively on pregnancy prevention (without aiming to prevent HIV) were excluded, as they may differ from programs with an HIV prevention component. Pregnancy-focused programs may have been less likely to accommodate the HIV-related risks of oral sex, anal sex, same-sex sexual behaviors, or non-sexual means of HIV transmission.

Types of outcome measures

Studies reporting outcome measures directly related to HIV transmission (i.e. ,self-reported risk behavior and biological outcomes) were included. Examples of risk behavior outcomes included condom use, number of sexual partners, and frequency of unprotected intercourse. Biological outcomes included the incidence of HIV, STIs, and pregnancy. If reports included a summary measure of sexual risk, authors were contacted for data on specific outcomes. Primary outcomes were those most indicative of HIV-risk behavior (although we recognized that pregnancy outcomes do not account for a number of HIV transmission mechanisms); by limiting these to biological outcomes only, we hoped to reduce the likelihood of achieving significance on multiple primary outcomes by chance. Where they were provided, we also examined outcome measures relevant to HIV knowledge, adverse outcomes, program fidelity, cost-effectiveness, and intervention acceptability.

Trials that did not report a behavioral or biological outcome measure were excluded.

Biological (primary) outcome measures

HIV incidence

STI incidence

Pregnancy incidence

Behavioral (secondary) outcome measures

Incidence and frequency of unprotected vaginal sex

Incidence and frequency of unprotected oral sex

Incidence and frequency of unprotected anal sex

Incidence and frequency of any vaginal sex

Incidence and frequency of any oral sex

Incidence and frequency of any anal sex

Number of sex partners

Use of male condoms

Use of female condoms

Abstaining from sex if condoms are not used

Duration of abstinence post-intervention

Return to abstinence (for those who were previously sexually active)

Incidence of sexual initiation (sexual debut)

Search methods for identification of studies

We refined our search strategy with recommendations from the Cochrane HIV/AIDS Review Group. No language restrictions were imposed and translations were sought where necessary. No restrictions on journal of publication were imposed, and no country names or other geographical terms were used in the search. Most databases were searched from 1980 onward. Databases were initially searched in May 2005, with all searches updated in February 2007. Full search strategies for each database are included in Table 1.

Electronic databases

We searched the following electronic databases, ending February 15, 2007:

1. ADOLEC (Inception-2007)
2. AIDSLINE (1980-2007)
3. AMED (1985-2007)
4. ASSIA (1987-2007)
5. BiblioMap (1887-2007)
6. BIOSIS (1969-2007)
7. BNI (1985-2007)
8. Catalog of US Government Publications (1976-2007)
9. CENTRAL (Cochrane Central Register of Controlled Trials) (1980-2007)
10. CHID (1985-2005; went offline in 2005)
11. CINAHL (1982-2007)
12. DARE (1991-2007)
13. Dissertation Abstracts International (1997-2007)
14. EMBASE (1974-2007)
15. ERIC (1991-2007)
16. EurasiaHealth Knowledge Multilingual Library (Inception-2007)
17. Global Health Abstracts (1973-2007)
18. HealthPromis (1997-2005; went offline in 2005)
19. HMIC (1983-2007)
20. PAIS (1972-2007)
21. Political Science Abstracts (1975-2007)
22. PsycINFO (1887-2007)
23. PubMed (1980-2007)
24. RCN (1985-1996; updating ended in 1996)
25. SCISEARCH (Web of Knowledge) (1974-2007)
26. SERFILE (Inception - 2005; inaccessible in 2007)
27. SIGLE (1980-2005; went offline in 2005)
28. Social Services Abstracts (1979-2007)
29. Sociological Abstracts (1963-2007)
30. TRoPHI (Inception-2007)

Other relevant libraries of international agencies, especially those concerned with the prevention of HIV/AIDS (UNAIDS, USAID, WHO, UNFPA, World Bank, and Centers for Disease Control and Prevention) were searched. We made additional efforts to identify and acquire unpublished literature.

Handsearching

Handsearching of various conference proceedings from 2000 onwards was conducted to identify unpublished reports. These included proceedings from the International AIDS Conference, the Conferences on Retroviruses and Opportunistic Infections, the US National HIV Prevention Conferences, the Abstinence Education Evaluation Conference, and the International Society of STD Research (ISSTDR).

Personal communication

We contacted leading experts in the field of abstinence-based programs to solicit potentially relevant unpublished papers, ongoing research, and suggestions for other contacts.

Cross-references

The reference lists of related reviews and primary studies were examined for additional citations.

Search terms

We searched with combinations of the following terms, truncating where possible. MeSH terms (e.g., "Sexual Abstinence") were also identified and included. We did not include country names or program names, as they may not be specified in searchable headings.

Intervention terms: abstinence, abstain, chastity, chaste, virgin, celibacy, celibate, sex education, marriage, delay, postpone.

Study terms: randomized controlled trial, controlled clinical trial, random allocation, double-blind, single-blind, clinical trial, mask, comparative study, control, pre-post controlled designs, comparison group, cohort study, comparative study, evaluation study, feasibility study, follow up studies.

HIV terms: HIV Infections, HIV, HIV-1, HIV-2, human immunodeficiency virus, human immunedeficiency virus, human immuno-deficiency virus, human immune-deficiency virus, AIDS, acquired immunodeficiency syndrome, acquired immunedeficiency syndrome, acquired immuno-deficiency syndrome, acquired immune-deficiency syndrome, sexually transmitted diseases.

PUBMED SAMPLE SEARCH STRATEGY

#1 Search HIV Infections[MeSH] OR HIV[MeSH] OR hiv[tw] OR hiv-1*[tw] OR hiv-2*[tw] OR hiv1[tw] OR hiv2[tw] OR hiv infect*[tw] OR human immunodeficiency virus[tw] OR human immunedeficiency virus[tw] OR human immuno-deficiency virus[tw] OR human immune-deficiency virus[tw] OR ((human immun*) AND (deficiency virus[tw])) OR acquired immunodeficiency syndrome[tw] OR acquired immunedeficiency syndrome[tw] OR acquired immuno-deficiency syndrome[tw] OR acquired immune-deficiency syndrome[tw] OR ((acquired immun*) AND (deficiency syndrome[tw])) OR "Sexually Transmitted Diseases, Viral"[MeSH:NoExp] Limits: Publication Date from 1980 to 2007

#2 Search SEXUAL ABSTINENCE [MESH] Limits: Publication Date from 1980 to 2007

#3 Search ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED)) Limits: Publication Date from 1980 to 2007

#4 Search #2 OR #3 Limits: Publication Date from 1980 to 2007

#5 Search randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh] OR clinical trial [pt] OR clinical trials [mh] OR ("clinical trial" [tw]) OR ((singl* [tw] OR doubl* [tw] OR trebl* [tw] OR tripl* [tw]) AND (mask* [tw] OR blind* [tw])) OR (placebos [mh] OR placebo* [tw] OR random* [tw] OR research design [mh:noexp] OR comparative study [mh] OR evaluation studies [mh] OR follow-up studies [mh] OR prospective studies [mh] OR control* [tw] OR prospectiv* [tw] OR volunteer* [tw]) NOT (animals [mh] NOT human [mh]) Limits: Publication Date from 1980 to 2007

#6 Search #1 AND #4 AND #5 Limits: Publication Date from 1980 to 2007

Data collection and analysis

Selection of studies

All reviewers independently examined citations reviewed by the search strategy. KU obtained full articles of potentially relevant papers. All reviewers independently assessed their eligibility using an eligibility form based on pre-specified inclusion criteria. Reasons for excluding potentially relevant trials are specified in the "Characteristics of Excluded Studies." We contacted the authors for clarification when necessary. Disagreements were resolved by discussion and, if necessary, referral to the Cochrane HIV/AIDS Review Group.

Data extraction and management

All reviewers independently extracted data on study design, participants, interventions, outcomes and methodological quality, using pre-designed data collection forms. Any disagreements were resolved by discussion.

The extracted data included the following measures: citation, study design, methodological criteria, inclusion and exclusion criteria, comparison group intervention, participant characteristics, trial setting, theoretical basis for intervention, elements of intervention, all relevant outcome measures, and results. We expected the programs and trial settings to vary widely (e.g., schools, clinics, prisons, community centers, and churches), and examined this carefully when considering the heterogeneity of our data set. We also examined any process or cost-effectiveness data reported, including training, monitoring, acceptability, costs, and sustainability.

Where reports were uncertain or included summary measures, authors were contacted for clarification. We noted any data that were consistently underreported, and highlighted this deficit along with future research needs in the discussion.

Assessment of methodological quality of included studies

Depending on study design, we evaluated the following methodological components, since there is evidence that these may be associated with biased estimates of effects:

1. similarity of the intervention and control groups (whether groups were treated equally except for the intervention);
2. method of generation of the randomization sequence;
3. generation of allocation concealment;
4. blinding (participants, intervention staff, and outcome assessors);
5. intention-to-treat analysis;
6. loss to follow-up. (In our results and discussion sections, we highlight attrition as a particular limitation of any studies with a total dropout exceeding one-third of baseline enrollment.)

Measures of treatment effect

To investigate the feasibility of a meta-analysis, all eligible studies were summarized in RevMan to the fullest extent possible. All reviewers assisted in abstracting the data, with KU entering all data into RevMan and DO and PM rechecking all entries. Disagreements were resolved by discussion. A narrative synthesis was provided for all results, but we determined that a statistical meta-analysis was not appropriate for this review due to differences in program and evaluation design, underreporting of key data, and inability to retrieve unpublished or missing data from authors. Had the studies been comparable and more completely reported, our meta-analysis would have measured the weighted average effect size for each outcome measure, weighted mean differences (for continuous outcomes), odds ratios of effects (for dichotomous outcomes), and 95% confidence intervals. Odds ratios less than one express a protective intervention effect.

Unit of analysis issues

Studies with multiple treatment groups

A meta-analysis was not conducted for this version of the review. However, had meta-analysis been possible, we would have dealt with multiple treatment groups in the following way: when studies had more than two treatment groups, we planned to accept only two groups for the meta-analysis. In keeping with procedures used by the US HIV/AIDS Prevention Research Synthesis project (Johnson 2002), we planned to take the arm with the greatest or most intense exposure (as identified by the trialist), compared to the arm with the least or least intense exposure, thereby giving each study the most favorable chance to yield significant program effects. When different trial arms emphasized sexual abstinence to a differing extent, we would have accepted the arm that emphasized abstinence the most, compared to the arm that emphasized abstinence the least.

Crossover trials

We did not anticipate dealing with these trials in the review, and no crossover trials were recovered by the search. If we had identified any studies using this design, we would have analyzed only the follow-up data from the period before the crossover.

Cluster-randomized trials

When studies use cluster randomization, analyzing the data at the level of the individual participant has two implications: first, p-values may be artificially small, suggesting significant effects where none occurred (Higgins 2005). Second, the confidence intervals for these results may appear artificially narrow, which would cause the studies to receive disproportionately more weight in the context of a meta-analysis (Donner 2002, Johnson 2002). The second concern was unimportant for this version of the review because no meta-analysis was conducted. However, the first remained relevant because wherever possible, we reported the individual study results as calculated in RevMan.

Before entering the results of cluster-randomized studies into RevMan, we transformed these data according to the procedure in the Cochrane Handbook (Higgins 2005, supported in Adams 2004), dividing the number of events and number of participants by the design effect [1 + (1 - m) * r]. We used the details provided by each study (total n and number of clusters) to calculate the av-

erage cluster size (m), and we decided upon consultation with the review group and a number of statistical experts to use the intra-class correlation coefficients (r) recommended in Johnson 2002. These were ICC=0.015 for school-based studies and ICC=0.005 for community-based studies (regardless of cluster size); these values were calculated respectively from school-based studies of youth smoking behavior and community-based studies of HIV prevention for men who have sex with men.

We acknowledge that we are uncertain about the applicability of these values to the context of this review. However, insufficient data were available for us to estimate a reliable ICC from the studies in this review. Since no meta-analysis was conducted, we report the results of cluster-randomized trials analyzed both before and after controlling for clustering. Insufficient data were available for us to calculate average cluster size for Anderson 1999, so these results are reported on an individual level.

Dealing with missing data

Missing data arose from two sources: participant attrition and missing statistics.

Attrition

To deal with participant dropout, we accepted studies that used any type of analysis, including complete case analyses (i.e., participants were analyzed according to original assignments, and no data were imputed for drop-outs). If a meta-analysis had been possible, we would have conducted sensitivity analyses to investigate attrition as a source of heterogeneity and possible bias.

Missing statistics

Where statistics were missing (e.g., numbers of participants per group, attrition rates, means and standard deviations, or percentages), we contacted primary study authors on two separate occasions to supply the information. Where the information was unavailable due to data loss or non-response, we reported the results as stated in the trial report.

Assessment of heterogeneity

If a meta-analysis had been possible for this version of the review, we would have assessed heterogeneity using the chi square test of heterogeneity, visual inspection of the data, and the I2 statistic (Higgins 2002). The I2 statistic determines the percentage of variability that is due to heterogeneity rather than sampling error, where a value greater than 50% suggests moderate heterogeneity. If any of these methods indicated heterogeneity, we would have investigated possible explanations, including clinical and methodological characteristics. Even when tests for heterogeneity were non-significant, we planned to conduct sub-group analyses to explore other potential moderators.

Assessment of reporting biases

If a meta-analysis had been possible for this version of the review, we would have used funnel plots (plotting effect size against standard error) to detect potential bias. Additional analyses may have included the planned Egger regression approach with a weight-function model. Asymmetry can be due to publication bias, but it can also be due to clinical and methodological heterogeneity. In the event that a relationship had been found, these sources of heterogeneity would also have been examined as possible explanations (Egger 1997).

Data synthesis

If a meta-analysis had been possible for this version of the review, we would have considered conducting analyses according to both fixed-effects and random-effects models. The random-effects model would have been used where there was indication of heterogeneity and the source of such heterogeneity could not be explained. The random-effects model would also have been used for analyses incorporating small numbers of studies, for which tests of heterogeneity may be underpowered. If there were no sources of heterogeneity beyond differences in the observed covariates, we would have conducted both fixed-effects and random-effects analyses and investigated differences between the two procedures.

The narrative synthesis was conducted according to the following methods. First, we entered each individual study in RevMan. Next, we prepared the supplemental Charts of Effects appended to this review (see Figures 1-10, Figure 1). These charts are organized by outcome, and each chart shows the effects of individual studies according to RevMan analyses (where possible) or as reported in the primary study. The charts also integrate key methodological features of each study (e.g., control group, attrition), along with each study's definition of the outcome of interest (e.g., condom use was variously defined as condom use at last sex among sexually experienced participants, or condom use in the last month among sexually active participants). The charts were then explained in text format for the description of studies and results.

Figure 1. Key to Figures 1-10.

KEY TO FIGURES	
Term	Definition
Abstinence-plus	An intervention that stresses abstinence as the best means of HIV prevention, but that also promotes condom use and other safe-sex strategies.
Adult-led	In Kirby 1997a, the experimental program was delivered by adult leaders.
Attrition	Overall attrition from the trial, expressed as the percentage of those randomized who dropped out of the trial at the given follow-up. This does not account for outcome-specific attrition due to nonresponse or reporting by subgroup.
Cluster-adjusted	Analyses control for clustering using the procedures outlined in the Cochrane Handbook, ICC=0.015 for school-based studies. ICC=0.005 would have been used for community-based studies.
Comparison	What treatment, if any, was delivered to the control group.
N	Number of participants entered into the analysis for the specific outcome.
Non-Enhanced	A lesser version of the experimental program (for Blake 2001, controls received the intervention program without parent-child activities; for Miller 1993, one control group received the intervention program without mailed newsletters).
Outcome measured	Trialist's definition of the outcome.
PI	The follow-up assessment was administered immediately after the intervention, but it was unclear how much time had elapsed since baseline.
Result	Outcome data, analyzed in RevMan where possible. All odds ratios (with 95% confidence intervals) were calculated in RevMan. All odds ratios less than one express a protective effect of the abstinence-only intervention. Odds ratios greater than one express an adverse effect. Where data could not be entered into RevMan, the results are reported as in the primary trials, with no further analysis by the reviewers. Red indicates a significant adverse effect (ie analyses favored controls over the abstinence-only program). Green indicates a significant protective effect (ie analyses favored the abstinence-only program over controls).
Safer-sex	An intervention that does not present abstinence as the best or exclusive preventive choice; preventive options (including abstinence) may be presented without a hierarchy in place, with a focus on safe-sex practices.
Time (mo)	Number of months that elapsed between baseline and the follow-up assessment.
Unadjusted for clustering	Analyses are conducted on the level of the individual, regardless of the unit of random assignment
Usual care	"Usual services" as defined by the trialist.
Youth-led	In Kirby 1997a, the experimental program was delivered by youth leaders.

Sub-group analyses

If we had conducted a meta-analysis, we would have stratified our analysis by intervention setting (e.g., school, clinic) and by participants' age, country, ethnicity, and socioeconomic status. Sub-group analyses would only have been conducted if data had allowed, with the knowledge that using a number of sub-groups can lead to statistically misleading conclusions.

Sensitivity analysis

If we had conducted a meta-analysis, we would have performed separate analyses for studies using complete case, per-protocol, and intention-to-treat methods, with a sensitivity analysis to investigate disparities between the three groups.

RESULTS

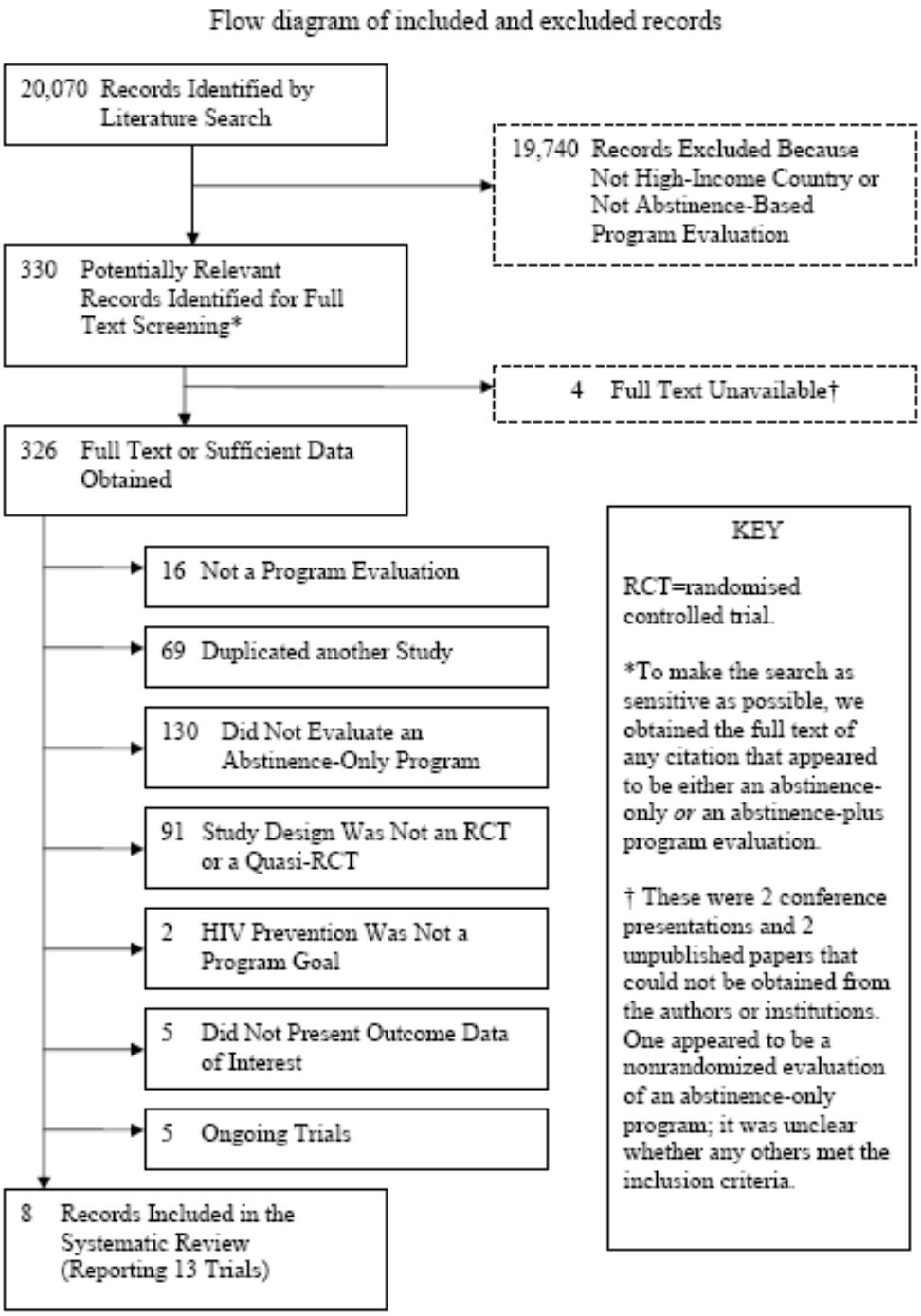
Description of studies

See: Characteristics of included studies; Characteristics of excluded studies; Characteristics of ongoing studies.

Results of the search

See Figure 2.

Figure 2. QUOROM Chart.



A total of 20070 citations were assessed for inclusion (of these, 19,892 were recovered from electronic database searches ending February 2007, 99 were from handsearching, 17 were from personal contacts, and 62 were from cross-referencing). After the screening process, 330 papers were deemed potentially relevant evaluation studies. Full-text copies (or sufficient information) for 326 of these citations were obtained.

Of the four papers we could not obtain, two were conference abstracts of unclear relevance whose authors could not be located. One was an in-press paper of unclear relevance that could not be obtained from the authors. One was no longer available in print and the authors could not be located; this appeared to be a non-random evaluation of an abstinence-only program.

Of the 326 papers for which sufficient study information was obtained, 318 were excluded for the following reasons: 16 were not program evaluations; 69 duplicated other papers; 130 did not evaluate an abstinence-only program; 91 evaluated an abstinence-only program but did not meet the criteria for study design; two evaluated abstinence-only programs using randomized controlled designs, but the programs did not list HIV prevention as a program goal (Jorgensen 1993, Mitchell-DiCenso 97); five evaluated an abstinence-only HIV prevention program and possibly used a randomized controlled design, but did not present outcomes of interest (Lage 2005, O'Donnell 2005, Pallone 2007, Roberts-Gray 2007, Tobin 2005), and five were ongoing trials (Borawski 2007, Jemmott 2006, Miller-Heyl 2007, Nagel 2007, Markham 2007). After these studies were excluded, analyses were limited to eight separate papers reporting 13 trials (Anderson 1999, Blake 2001, Clark 2005, Goldfarb 1999, Hernandez 1990, Kirby 1997a, Kirby 1997b, Kirby 1997c, Miller 1993, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). One paper described three evaluations (Kirby 1997a, Kirby 1997b, Kirby 1997c), and two supplementary papers duplicated these reports (Kirby 1995, Cagampang 1997). Another paper described four evaluations (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), and was also supplemented by two reports (Devaney 2002, Maynard 2005). One paper was included on the basis of unpublished information about the method of allocation (Goldfarb 1999).

Included studies

Detailed information about individual studies may be found in the Table of Included Studies.

Design

All of the abstinence-only evaluations were randomized controlled trials; no quasi-randomized controlled trials evaluating abstinence-only programs were retrieved by the search. Of the 13 evaluations included, three were part of the same report (Kirby 1997a, Kirby 1997b, Kirby 1997c), which used separate evaluation designs to assess Postponing Sexual Involvement (PSI) program. These differed by units of randomization (which were classroom, school, and individual, respectively). Four evaluations were included in another report (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d); these differed by site and experimental program.

Six of the evaluations involved cluster randomization. The unit of randomization was the school for two evaluations (Goldfarb 1999, Kirby 1997b), the classroom for three evaluations (Blake 2001, Clark 2005, Kirby 1997a), the individual for six evaluations (Hernandez 1990, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), the family for one evaluation (Miller 1993 - it appeared that only one "target teen" per family was included), and "naturally occurring groups" in the final evaluation (Anderson 1999; this designation included the school unit).

Control groups varied. One study compared an abstinence-only program to a non-intervention control (Anderson 1999); the control group also received the intervention after the 12-month follow-up assessment.

Blake 2001 compared an enhanced version of a program against a non-enhanced version of the same program. The control group was exposed to the Managing the Pressures before Marriage (MPM) curriculum only, while the other received the curriculum plus extra homework assignments to facilitate parental involvement.

Hernandez 1990 involved four arms: abstinence-only (i.e., "Abstinence," which presented abstinence as the best and only strategy for avoiding HIV), abstinence-plus (i.e., "Protection," which presented abstinence as the safest alternative and safer sex as another alternative, focusing on condoms), safer-sex (i.e., "Decision," which focused on safer sex and personal decision-making), and a non-intervention control. University students were randomized to each of the four arms. All programs took place in the same setting and used the same one-session delivery format.

Miller 1993 used a three-arm design to assess the effect of a video-based intervention intended to enhance parent-child communication about sex and abstinence. Families were randomized to receive one of three programs: videos plus mailed newsletters suggesting further family discussions, videos alone, and no intervention.

Nine effectiveness studies compared an abstinence-only program against usual care (Clark 2005, Goldfarb 1999, Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), with "usual care" defined as the HIV-prevention intervention normally delivered by the school. Since few details were provided to describe what "usual care" consisted of, it was impossible to clarify whether these control groups received safer-sex programs, abstinence-plus programs, other abstinence-only interventions, or any combination of the above. The available details were as follows:

An unpublished report for Kirby 1997a, Kirby 1997b, and Kirby 1997c states that "the vast majority of students in the control

groups did not receive another sexuality education curriculum instead of PSI. Instead, they received instruction in some other topic area ... many students in both the PSI groups and the control groups did receive some type of instruction on sexuality either before or during the time period of this evaluation." Although the published report of Kirby 1997c states that the study used a no treatment control, the unpublished evaluation states that control group participants received the "standard curriculum" delivered by the community agencies.

The report for Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, and Trenholm 2007d provides some detail about "usual care" in each trial. For Trenholm 2007a, controls received a nine-week physical education class that did not discuss abstinence or HIV, as well as a health class emphasizing abstinence but providing no details about HIV. Controls in Trenholm 2007b received a week-long unit on human growth and development, as well as a curriculum covering HIV and abstinence. Controls in Trenholm 2007c received "units on abstinence and contraceptive use" beginning in fifth grade. Finally, controls in Trenholm 2007d received "heath, family-life, and sex education curricula," but it was unclear what these classes entailed.

Sample sizes

At baseline, the studies together enrolled approximately 15940 participants. Sample sizes for the abstinence-only evaluations at the time of baseline assessment ranged from 248 (Clark 2005) to 5244 (Kirby 1997b); baseline enrolment was not specifically reported and could not be computed for Goldfarb 1999. The median baseline enrolment was approximately 551. Five trials reported using a power calculation (Clark 2005, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d).

After attrition at each study's longest follow-up, the analyses reported in this review encompassed approximately 11879 participants. Again, this value is somewhat uncertain due to unclear reporting in several of the primary studies, and it is an overall estimate (i.e., it does not account for non-response on any specific outcome measure).

Setting

All 13 trials enrolled adolescents in the United States. Program sites included California (Kirby 1997a, Kirby 1997b, Kirby 1997c, Anderson 1999), upstate New York (Blake 2001), "a southeastern metropolitan area" (Clark 2005), New Jersey (Goldfarb 1999), North Carolina (Hernandez 1990), Utah (Miller 1993), Virginia (Trenholm 2007a), Florida (Trenholm 2007b), Wisconsin (Trenholm 2007c), and Mississippi (Trenholm 2007d).

The interventions were delivered in elementary and/or middle schools (Blake 2001, Clark 2005, Goldfarb 1999, Kirby 1997a, Kirby 1997b, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d); a university (Hernandez 1990); community centers (including churches), youth recreation programs, and social service agencies (Kirby 1997c); a mixture of schools and community centers (Anderson 1999); and family homes (Miller 1993).

Participants

All participants were adolescents or young adults in the US. All but one study (Hernandez 1990) targeted youth between ages 9-13 years; the remaining study targeted young people aged 18 to 21 years. The mean participant ages in the 13 studies ranged from 10.6 years (Anderson 1999) to 19.25 years (Hernandez 1990), the median of these average ages was 12.8 years. Participant groups in all but one study included both males and females; one trial included females only (Trenholm 2007b).

Participants were primarily African-American in four studies (Clark 2005, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), largely of minority ethnicity in three studies (Anderson 1999, Kirby 1997a, Kirby 1997c), "high minority" in another study (Goldfarb 1999), equally representative of white and minority ethnicity in one study (Kirby 1997b), and primarily white in four studies (Blake 2001, Hernandez 1990, Miller 1993, Trenholm 2007a). One study took place in a community with a high proportion of Mormon families, who were described as having a "highly abstinence-valued culture" (Miller 1993).

Four studies reported little or no information about the socioeconomic status of participants (Hernandez 1990, Kirby 1997a, Kirby 1997b, Kirby 1997c), although the Kirby studies took place in areas with elevated teen pregnancy rates, potentially indicating lower socioeconomic status. The communities of two studies were described as "middle-class" (Blake 2001, Trenholm 2007a), and five were of lower socioeconomic status (Clark 2005, Goldfarb 1999, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). One evaluation mentioned that participants' mothers' education varied between high school graduation and some college coursework (Anderson 1999). In the family-based study, 38% of mothers and 56% of fathers were college-educated (Miller 1993).

Parental consent procedures were relevant ethical concerns for this review, since 12 of the studies involved legal minors as participants. Eight studies used active parental consent procedures (Anderson 1999, Blake 2001, Clark 2005, Miller 1993, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d); three studies required parental notification, with permission forms used by 54% of the contractors hired to deliver the intervention (Kirby 1997a, Kirby 1997b, Kirby 1997c); and one study used passive parental consent procedures (Goldfarb 1999). Rates of parental opposition in every study were low. Incentives for participation were offered in one study that granted its university-level participants course credits (Hernandez 1990), one study that gave participants prizes for homework completion (Blake 2001), one study that gave participants $5 for each survey completed (Clark 2005), and three studies for which contractors often had "reunions" or pizza parties to facilitate the administration of post-test surveys (Kirby 1997a, Kirby 1997b, Kirby 1997c).

Interventions

The 13 abstinence-only programs evaluated in this review had many similarities. All of the programs stressed sexual abstinence as the best and only means of HIV prevention, and no program

promoted condom or contraceptive use as a viable alternative to abstinence. Eleven of the programs were at least partially school-based (Kirby 1997c and Miller 1993 were the exceptions), and most of the programs emphasized communication and decision-making. Five of the interventions were based on the same program model: Kirby 1997a, Kirby 1997b, Kirby 1997c, and Trenholm 2007d assessed the effects of Postponing Sexual Involvement (PSI), which was modified slightly to form the curriculum Managing the Pressures before Marriage (MPM) as evaluated in Blake 2001. All but one of the programs were designed for adolescents and delivered to middle-school-aged youth. Trenholm 2007d also involved components of the Sex Can Wait program, which was evaluated in the study by Goldfarb, et al. (Goldfarb 1999).

Although the interventions did appear similar enough to permit a systematic review, the information reported about intervention design and implementation was variable and generally limited. Although most reports specified basic information about intended delivery format, subject material, activities, and exposure, very few collected or reported data on adherence, implementation fidelity, actual exposure, context, or modifications to the original design. Additionally, it was often difficult or impossible to quantify the extent to which the subject matter of a program focused on abstinence (and excluded information regarding contraception and condom use), which made the recognition of abstinence-only programs somewhat difficult. There is a need for more detailed information on intervention design and actual implementation in primary trial reports.

Importantly, 12 of the interventions emphasized pregnancy prevention to an equal or greater extent than HIV prevention. One experimental program was limited to HIV prevention (Hernandez 1990). It is unclear how this evidence generalizes to programs with an exclusive HIV prevention focus.

All but one of the evaluations (Hernandez 1990) involved multiple sessions, and every program aimed to stop, decrease, or postpone sexual activity within the participant population. Six programs appeared to be delivered exclusively in a large-group classroom format (Blake 2001, Goldfarb 1999, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). These programs may also have utilized small-group activities, but this was not stated explicitly. One of these interventions (Blake 2001) supplemented classroom activities with parent-child homework for completion in the family home. Trial reports for five programs mentioned both classroom and small-group activities (Anderson 1999, Clark 2005, Kirby 1997a, Kirby 1997b, Kirby 1997c). Two programs were media-based: one included videos and pamphlets delivered in a university classroom (Hernandez 1990), while the other included videos viewed in the family home (Miller 1993).

Trial reports indicated that at least eleven of the interventions aimed to improve communication between parents and adolescents (Anderson 1999, Blake 2001, Goldfarb 1999, Miller 1993, Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). Strategies for parent involvement included homework assignments (Anderson 1999, Blake 2001, Goldfarb 1999), parent meetings or "parent nights" (Anderson 1999, Kirby 1997a, Kirby 1997b, Kirby 1997c), home visits by social workers (Trenholm 2007b), a parent curriculum (Trenholm 2007a), and videos for parental viewing (Miller 1993). However, reports acknowledged that parental uptake of these activities was often low. Parent-child activities were not mentioned in two studies (Clark 2005, Hernandez 1990).

Program exposure (i.e. dosage) varied from one session (Hernandez 1990) to 720 sessions (Trenholm 2007c), with a median exposure of 8 sessions across trials.

Staff members for most interventions (Anderson 1999, Clark 2005, Goldfarb 1999, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d) were adults, including teachers and public health personnel. One design randomized classrooms to peer-led PSI, adult-led PSI, or a control group (Kirby 1997a), while another intervention was delivered entirely by peer leaders (Blake 2001). The two media-based interventions had minimal staff involvement by adults (Hernandez 1990, Miller 1993).

Four of the evaluation reports stated theoretical bases for the interventions; these included cognitive behavioral theory (Anderson 1999), social learning theory (Anderson 1999, Blake 2001, Goldfarb 1999), social cognitive theory (Blake 2001), and the theory of possible selves (Clark 2005). Four additional evaluations used a program based on social inoculation theory (Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007d, original PSI program design reported in Howard 1990 and Howard 1992).

Specific intervention details are as follows:

Anderson 1999 evaluated Reaching Adolescents and Parents (RAP), which targeted students in grades 5-7 and their parents. The eight-session curriculum focused on puberty, self-esteem, communication, relationships, and decision-making. Six sessions involved the youth alone; one involved joint parent-child activities focusing on communication (e.g. discussions, video viewing); and the final session was a discussion forum for parents.

Blake 2001 assessed the effects of Managing the Pressures before Marriage (MPM), a five-session curriculum for middle school students modeled on PSI, but with an additional emphasis on delaying sex until marriage. The five sessions were delivered by slightly older peer leaders (i.e., high-school aged) in a classroom setting. The independent variable in this evaluation was the addition of five parent-child homework assignments to the curriculum; these assignments aimed to improve parent-child communication and focused on concepts such as personal values, HIV-related myths, reasons to postpone sex, sex in the media, self-esteem, and peer pressure.

Clark 2005 evaluated the Adult Identity Monitoring program, a 10-session program delivered by a male-female pair of African-American adult facilitators. The program targeted African-American adolescents, and it attempted to change behavior by helping participants to envision their future selves and understand how

risk-taking might affect their goals.

Goldfarb 1999 evaluated Project C.A.R.E. (Creating Awareness through Relationship Education), a classroom initiative with the primary focus of delaying sexual activity. The curriculum focused on communication, factual information about sexual behavior, the consequences of sex, goal-setting, social support, and decision-making. Participants in this study received slightly different programs depending on their grade levels: sixth grade students received 23 sessions based on the "upper elementary" Sex Can Wait curriculum, seventh grade students received 24 sessions from the "middle school" Sex Can Wait curriculum, and eighth grade students received 24 of the 45 sessions in the Abstinence Pick and Choose Activities curriculum. All participants were grouped together for analysis, which made it impossible to isolate the effects of each program. It was unclear how many students were in each grade level.

Hernandez 1990, the only intervention to take place in a university setting, compared three active intervention groups to a non-intervention control. In a large group setting, participants in each arm received a separate seven-page health education curriculum with information about the biology, symptoms, and transmission of HIV. Participants then viewed a 20-minute AIDS prevention film that included commentary and interviews; each intervention arm watched a separate video. Lastly, participants in each arm received a separate brochure highlighting transmission modes, symptoms, and prevention of HIV. The content of the three programs varied, with the intent of following three "prototypes of programs advocated by boards of education, health organizations, or health educators." One program focused on abstinence only ("Abstinence"), one prioritized abstinence but also promoted safer sex behaviors ("Protection"), and one focused on safer sex ("Decision"). This was the only study to explicitly compare an abstinence-only program against an abstinence-plus or a safer-sex design.

Kirby 1997a, Kirby 1997b, and Kirby 1997c assessed the effects of Postponing Sexual Involvement (PSI), a five-session curriculum for middle school students focusing on the risks of early sex, reasons to delay sex, resisting social pressures, limit-setting, and communication. This program was implemented as part of the California initiative ENABL (Education Now and Babies Later). Kirby 1997a systematically varied the staff delivering the intervention between adult health educators and peer leaders, comparing each to a usual care control group. Kirby 1997b and Kirby 1997c each used adult staff to deliver the intervention. The PSI curriculum has been evaluated elsewhere as an abstinence-plus program with the addition of a curriculum including contraception instruction (Aarons 2000, Howard 1990, Howard 1992), but it functioned as an abstinence-only intervention in this study (Kirby 2002).

The one intervention not delivered in a group setting was Facts and Feelings, a video-based curriculum designed for weekly viewing by parents and adolescents in the home (Miller 1993). Six 15- to 20-minute videos were distributed to parents for the duration of three months, with biweekly follow-up calls by intervention staff to collect usage data. Videos focused on puberty, sexual values, facts about sexual anatomy and reproduction, the media's portrayal of sex, the consequences of sex, and refusal skills. The study compared two active treatment groups to non-intervention controls; one treatment group received the videos along with supplementary newsletters designed to suggest further family discussions and discussions, while the other group received the videos only.

Trenholm 2007a compared a mandatory school-based abstinence-only program to usual care. The program, entitled My Choice, My Future! was a composite of three curricula: Reasonable Reasons to Wait (30 classroom sessions over the eighth-grade year focusing on character development, relationships, and abstinence), The Art of Loving Well (eight classroom sessions over the ninth-grade year focusing on healthy relationships), and WAIT Training (14 sessions over the tenth-grade year focusing on relationships skills, risks, and a slide show about STDs and abstinence).

Trenholm 2007b assessed an elective program for 8th-grade girls entitled "ReCapturing the Vision," comparing the program to usual care. The program met daily throughout one school year (180 sessions), and was a composite of two programs. "ReCapturing the Vision" focused on personal and career goals and included home visits by social workers, tutoring, community service, a retreat, a Teen Abstinence Rally, and a symposium with celebrity speakers. "Vessels of Honor" was added to this curriculum; this program focuses on "honorable behavior," communication, relationships, physical development, sexual abuse, and "strategies for choosing a mate and the benefits of a committed marital relationship." Once girls elected to join the program, attendance was mandatory.

Trenholm 2007c assessed an elective program entitled "A Life Options Model Curriculum for Youth," which met for 2.5 hours daily during the school year, for four consecutive years (approximately 720 program sessions). The program focused on self-esteem, values, decision-making, risks, communication, relationships, sexuality, development, sexually transmitted infections, and social skills. The program also included parent workshops and mentoring programs. Attendance was voluntary.

Trenholm 2007d compared the Teens in Control program to usual care. Teens in Control was a composite of Postponing Sexual Involvement (also evaluated in Kirby 1997a, Kirby 1997b, and Kirby 1997c) and Sex Can Wait (also evaluated in Goldfarb 1999). The program met for two hours weekly throughout one school year (approximately 36 program sessions), and focused on risks, pressures, puberty, self-esteem, values, communication, career goals, and skills to facilitate abstinence.

Outcomes

All outcomes were assessed via written self-report questionnaires, which were bilingual (English/Spanish) in four of the evaluations (Anderson 1999, Kirby 1997a, Kirby 1997b, Kirby 1997c). Questionnaires were generally completed in the intervention setting and administered to groups (or, in Miller 1993, to parents

and children separately) by project staff or intervention facilitators. Questionnaires in four trials were sometimes administered by phone (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). One report mentioned modifying questionnaires in response to opposition from concerned parents by removing questions from the survey (Anderson 1999). Nine of the evaluations described piloting the surveys (Anderson 1999, Hernandez 1990, Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d); three used or slightly modified surveys developed by others, but did not explicitly describe piloting procedures (Clark 2005, Goldfarb 1999, Miller 1993), and one did not explicitly report how the survey was developed (Blake 2001). No evaluations confirmed the results of questionnaires with medical records or biological assessments, making all data vulnerable to recall bias and the limitations of self-report.

The studies used a variety of techniques to improve the validity of self-reported data. These included having bilingual (English/Spanish) staff assist students during the interviews (Anderson 1999), reading questionnaires aloud (Anderson 1999, Hernandez 1990), allowing students to fill in the questionnaires anonymously (Anderson 1999, Goldfarb 1999, Hernandez 1990), emphasizing the confidential and voluntary nature of the exam before questionnaire administration (Blake 2001), seating participants far apart during survey administration (Hernandez 1990, Kirby 1997a, Kirby 1997b, Kirby 1997c), asking participants to sign consent forms that highlight confidentiality (Hernandez 1990), monitoring the survey room to prevent talking or participant wandering (Hernandez 1990, Kirby 1997a, Kirby 1997b, Kirby 1997c), using ID numbers and cover sheets to protect participants' identities (Kirby 1997a, Kirby 1997b, Kirby 1997c), not placing any personal identifying information on any survey form or data file (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), giving students papers to cover their answers (Kirby 1997a, Kirby 1997b, Kirby 1997c), and placing surveys in unmarked envelopes (Miller 1993).

Data were generally reported as statistical tests and/or group percentages (with or without the numbers of participants per trial arm); no study reported effect sizes. One study provided statements of significance without group means or other specific data (Miller 1993).

Timing of follow-up assessments ranged from six weeks (Hernandez 1990) to approximately 65 months (Trenholm 2007b) after baseline. For the purposes of this review, follow-up data were blocked into short-term (0-5 months after baseline), medium-term (6 through 11 months), and long-term (12 months or more). Ten of the 13 studies reported long-term follow-up data (Anderson 1999, Clark 2005, Kirby 1997a, Kirby 1997b, Kirby 1997c, Miller 1993, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d); the other three were limited to short-term follow-ups (Blake 2001, Goldfarb 1999, Hernandez 1990). The median final follow-up assessment occurred 17 months after baseline.

The recall periods for self-reported outcomes varied. Outcomes such as STI, pregnancy, number of partners, and sexual initiation tended to be reported for participants' lifetimes, while recall periods for outcomes such as recent vaginal sex or condom use included 1 month (Goldfarb 1999), 1.5 months (Hernandez 1990), 3 months (Blake 2001) 12 months (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), time since baseline (Clark 2005, Kirby 1997a, Kirby 1997b, Kirby 1997c), and ever (Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d).

The biological outcomes of interest for this review were underutilized. No study evaluated HIV incidence, and seven evaluations, conducted by two groups of trialists, assessed the incidence of STI. Eight studies assessed pregnancy incidence.

The behavioral outcomes of most interest (i.e., unprotected oral, anal, and vaginal sex) were also underutilized. Five studies assessed the incidence of unprotected vaginal intercourse among all participants. No evaluation separated sexual behavior outcomes into vaginal, oral, and anal behaviors; we entered the non-specific outcomes "intercourse" and "sex" as vaginal sex. Seven studies reported incidence of vaginal sex. Four studies evaluated the frequency of recent vaginal sex, and eight evaluated number of partners . Condom use and sexual initiation were reported in eight and ten trials respectively; these were the most commonly reported behavioral outcomes.

One study (Miller 1993) created a sexual behavior index for outcome observation, asking participants to report lifetime behaviors ranging from "holding hands" to "the sexual act by which pregnancy can occur." Data on vaginal intercourse could not be disaggregated from this index, making the result less relevant for this review.

Excluded studies

Excluded studies consist of all evaluations produced by the search that evaluated abstinence-only programs using non-randomized controlled, quantitative designs among any population group in high-income countries. Cross-sectional studies and comparisons of program participants against national survey samples were included. These abstinence-only programs could focus on preventing pregnancy or preventing HIV, and the studies could report any outcomes. Details of the excluded studies, including results for behavioral and biological outcomes only, are provided in the Table of Excluded Studies. The results mentioned in this table are those reported by the trialists, with no further analysis by the reviewers.

Ongoing studies

Information regarding ongoing studies (Jemmott 2006, Markham 2007, Miller-Heyl 2007, Nagel 2007, Borawski 2007) is provided in the Table of Ongoing Studies.

Risk of bias in included studies

Allocation

Four trials reported the method of randomization (via a random

number generator, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). One study did not report in the published version that it was a randomized controlled trial, but this was discovered upon communication with the trialist (Goldfarb 1999). Overall, the potential for inadequate randomization and allocation concealment is a severe limitation of this data set.

Randomization was made particularly difficult by the need to contain the potential for contamination; this led Anderson 1999, Blake 2001, Clark 2005, Goldfarb 1999, Kirby 1997a, and Kirby 1997b to use cluster randomization, which heightened the potential for baseline differences. Seven trials reported at least one significant baseline difference (Blake 2001, Clark 2005, Kirby 1997b, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), and each of these studies statistically controlled for these differences in analyses. One study (Miller 1993) did not provide an explicit statement of group equivalence at baseline.

Contamination and threats to program differentiation were openly acknowledged in the three studies that evaluated the same program (PSI) in California (Kirby 1997a, Kirby 1997b, Kirby 1997c). This three-study design may have allowed control-group participants for one evaluation to receive the intervention in another setting (e.g., a control participant in a school-based study could have received the intervention as an experimental participant in the community-based study). Threats to program differentiation could also have occurred in Blake 2001, for which the same personnel delivered both the experimental and control interventions. This study monitored program leaders to assess program differentiation and the potential for performance bias. The potential for contamination was unclear in four studies (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d), as it was not stated who delivered the experimental and control programs.

Allocation concealment was not specifically addressed by any study, suggesting that every study may be prone to selection bias. We gave every study the "B" designation for quality of allocation concealment, indicating that procedures for allocation concealment were not reported.

Blinding

Given the time and human contact required for these interventions, the blinding of participants and intervention staff was generally not feasible. The impossibility of blinding intervention staff may have led to performance bias.

The blinding of assessors and data collectors was not entirely crucial for this group of studies, because every study utilized written self-reported questionnaires for data collection. As mentioned, questionnaires were administered in group settings by trained data collectors; blinding of these data collectors was not reported in any study. Notably, Goldfarb 1999 used students' own teachers as monitors for data collection, which may have affected self-report if students had confidentiality concerns.

Follow-up and exclusions

Attrition was not known for one study, which did not report the number of participants randomized at baseline (Goldfarb 1999). Final attrition rates in the other nine studies ranged from 5% (Hernandez 1990) to 45% (Kirby 1997c), with a median final attrition rate of 20%. Four studies had a final attrition above 33%, which must be considered when judging both internal and external validity; these included the following: Anderson 1999 (38.0%), Clark 2005 (37%), Kirby 1997a (33.7%), and Kirby 1997c (45.0%).

Five studies reported attrition analyses (Anderson 1999, Clark 2005, Kirby 1997a, Kirby 1997b, Kirby 1997c). In Anderson 1999, participants who dropped out of the program were older by one year, more likely to be male, and less likely to report skipping school than the participants retained. Kirby 1997a, Kirby 1997b, and Kirby 1997c reported attrition analyses for the entire three-study design; approximately 1% more participants were lost from the intervention than the control group. Using two-way ANOVAs, the study found no significant interaction effects between treatment group and remaining in the study (i.e., survey completion at 17-month follow-up) in relation to age, gender, family background, risk factors such as alcohol use or low grades, mediating variables, or outcome variables. In Clark 2005, there was no difference in attrition by group; this report further specified that there were no differences by age, sexual initiation, sexual activity, birth control, or condom use.

Attrition analyses were not available for Blake 2001, but the study reported that non-completion of homework assignments in the MPM-enhanced group was significantly higher among males, students reporting recent sexual intercourse, students who did not receive "mostly A's in school," and black or Hispanic students (as opposed to non-Hispanic whites). Goldfarb 1999 also did not provide attrition analyses; however, although groups were reportedly comparable on all variables at baseline, follow-up comparisons on demographic variables showed several differences. Fewer control group participants were in the oldest age group or highest grade in school, and there were significant differences in living arrangements (the majority of control group participants lived in an apartment, while the majority of intervention participants lived in a house).

Hernandez 1990 and Miller 1993 did not report attrition analyses. The studies by Trenholm et al (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d) also did not report attrition analyses; attrition was slightly higher in the control group in three trials (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c), and slightly higher in the intervention group in one trial (Trenholm 2007d).

All 13 evaluations employed complete case analyses (Hollis 1999), although one study had 0% attrition at short-term follow-up, allowing analyses for that follow-up to follow intention-to-treat procedures (Anderson 1999). Since several studies reported that dropouts may have been at higher risk than those who were assessed at follow-up; the use of a complete case analysis may obscure intervention effects for the most at-risk participants.

Selective reporting

The quality of outcome reporting varied across studies, as did the statistical methods of analysis and the sub-groups for which data was reported. The variable reporting among the 13 trials suggests that no standard set of outcomes exists for the evaluation of sexual behavior for this population.

One trial collected data on number of sex partners and regularity of condom use, but did not report these data due to the small size of the subset of participants reporting sex (Blake 2001).

One trial originally collected data on frequency of sex, but the item was removed from the survey at follow-up due to parental objections, resulting in de facto selective reporting of a pre-defined outcome (Anderson 1999).

One study aggregated data into a sexual behavior index (Miller 1993), but did not disaggregate individual risk behaviors; this resulted in the data being less useful for the evaluation of HIV-related risk.

A common source of potential confusion was the use of non-specific self-report outcome measures, such as "sexual intercourse," "sex," or "virginity." Given that these terms may be difficult to understand, or may have many different definitions among adolescents, adults, medical practitioners, program coordinators, and the general public (Haglund 2002, Horan 1998, Pitts 2001, Remez 2000, Sanders 1999, Sonfield 2001), these constructs may require more specific definitions. No study separated oral, anal, and vaginal behaviors.

Some results, particularly those for outcomes such as condom use or number of partners, were analyzed only among participants with certain levels of sexual experience. Eight studies made these distinctions (Goldfarb 1999, Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). Goldfarb 1999 limited condom use to participants reporting sex (either in the past month or the past year, although this was unclear), while Kirby 1997a, Kirby 1997b, and Kirby 1997c limited analyses for frequency of vaginal sex, number of partners and condom use to participants reporting sexual experience at baseline or follow-up. Clark 2005, Kirby 1997a, Kirby 1997b, and Kirby 1997c appropriately assessed sexual initiation among participants who reported never having had sex at baseline. The trials by Trenholm, et al (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d) only assessed STI, pregnancy, and condom use among participants reporting sex ever. All other outcomes were assessed among all participants.

The data were reported in a variety of formats, many of which were unsuitable for meta-analysis in RevMan due to missing data. We made repeated attempts to contact trialists for missing statistical data; while many authors were helpful in providing additional data, further information was often unavailable due to data loss or non-response.

Blake 2001, Clark 2005, Goldfarb 1999, Anderson 1999, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, and Trenholm 2007d provided sufficient data for us to analyze all relevant outcomes in RevMan. We had insufficient data to control for cluster randomization for Anderson 1999. This study evaluated pregnancy and observed only one event in each trial arm at 12-month follow-up (a non-significant effect), so controlling for cluster randomization was unlikely to change the results. Kirby 1997a, Kirby 1997b, and Kirby 1997c provided sufficient information to enter dichotomous outcome data, but we could not enter continuous outcomes because we lacked standard deviations for group means. We had sufficient information to dichotomize outcomes for number of partners for four studies (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d); however, we are unaware of any consensus on a clinically meaningful difference (e.g., did or did not have more than three lifetime sex partners), so we have accepted the trialists' analyses according to frequency distribution.

Although controlling for baseline values was a strength of these reports, some of the statistical analyses reported did not control for the unit of randomization or correct for multiple tests. Of the six studies that used cluster randomization (Anderson 1999, Blake 2001, Clark 2005, Goldfarb 1999, Kirby 1997a, Kirby 1997b), two performed analyses on an individual level without reporting statistical procedures to accommodate the unit of assignment (Anderson 1999 and Goldfarb 1999). The other four studies incorporated multi-level analyses to control for cluster randomization (Blake 2001, Clark 2005, Kirby 1997a, Kirby 1997b). Only three studies used a correction to accommodate multiple statistical tests, accepting a significance level of $p<0.01$ instead of $p<0.05$ (Kirby 1997a, Kirby 1997b, Kirby 1997c). The lack of control for multiple statistical tests makes the remaining papers more vulnerable to Type I error (i.e., believing that programs significantly affected behavior, when significance occurred due to chance).

Other potential sources of bias

The reliance on self-reported data is a salient but inevitable source of bias for all studies evaluating sexual behavior. The limitations of self-report for sexual behavior have been analyzed elsewhere (Alexander 1993, Binson 1998, Brener 2004, Catania 1996, Lauritsen 1997, Meston 1998, Newcomer 1988, Rosenbaum 2006, Zabin 1984), particularly with respect to adolescents, and these data are also vulnerable to misreport due to the variety of definitions for terms such as "sex" or "partner." Self-report biases may be heightened in evaluations of abstinence-only programs due to the nature of the interventions; recent survey data suggest that youth who take virginity pledges may be more likely to retract their sexual histories at a later date (Rosenbaum 2006). Furthermore, there are demonstrable limits to the use of sexual behavior as a proxy for HIV/STI risk (Peterman 2000). The most robust data for abstinence-only program evaluations would be medical evaluations of HIV or STI incidence; however, due perhaps to the young age of the participants, resource limitations, or logistical complications, no study was able to evaluate outcomes in this way. Due to the limited space allocated to program design and implementation in published journals, the majority of studies reported little data on these aspects of the evaluation. A lack of specific

definitions of "abstinence" and "sex" leaves programs vulnerable to varying interpretations by participants and staff. Different programs may emphasize abstinence to differing extents or provide different amounts of information regarding the limitations of condom use; however, we could not uncover these differences from the available data. Some of our difficulty identifying abstinence-only programs and program similarities may be partly due to the controversial political climate surrounding such programs, along with confusion regarding the difference between abstinence-only and abstinence-plus designs.

All 13 of the evaluations attempted to some extent to monitor program dosage, promote treatment adherence, or assess the quality of implementation, but few details regarding actual implementation were provided. Efforts to promote implementation fidelity included monitoring attendance, homework completion, or video exposure (Blake 2001, Kirby 1997a, Kirby 1997b, Kirby 1997c Miller 1993); making trial staff available to assist in the conduct of program sessions (Blake 2001); having independent evaluators observe program delivery (Anderson 1999); requiring quarterly implementation reports from program facilitators (Kirby 1997a, Kirby 1997b, Kirby 1997c); interviewing program facilitators, program contractors, and participants (Kirby 1997a, Kirby 1997b, Kirby 1997c); conducting site visits and observing program sessions (Kirby 1997a, Kirby 1997b, Kirby 1997c); using standard intervention materials; and using treatment manuals. Seven studies reported to any extent on training for program facilitators (Anderson 1999, Blake 2001, Clark 2005, Goldfarb 1999, Kirby 1997a, Kirby 1997b, Kirby 1997c). Two programs were media-based and used standard materials (Hernandez 1990, Miller 1993). It is unclear whether strategies for monitoring and enhancing adherence were successful; few studies provided details on the actual extent of implementation fidelity (e.g., proportion of core components actually delivered, participant attendance rates).

In the trials by Trenholm, et al, attendance data suggested that the two elective programs (Trenholm 2007b, Trenholm 2007c) were poorly attended. In Trenholm 2007b, 35% of participants assigned to the experimental group chose not to participate; however, they were appropriately analyzed according to original assignment. In Trenholm 2007c, 43% of these participants chose not to participate in any of the program activities. Of the 57% experimental-group participants who did participate in Trenholm 2007c, 55% participated in less than half of the daily program sessions available in the first year (a total of approximately 720 daily program sessions were offered over the course of four years).

Effects of interventions

The protocol for this review specified that a meta-analysis would be conducted if appropriate. After a thorough investigation of the data set, we determined that a statistical synthesis of the given data would be uninformative for several reasons.

Incomparability of interventions: One compelling reason to refrain from meta-analyzing the results was clinical heterogeneity; there were differences among the 13 programs in terms of intervention design, delivery, exposure, staff, and extent of emphasis on abstinence. These differences made the programs unsuitable for aggregation, since a meta-analysis aggregating disparate interventions may not provide useful insight regarding which types of programs are effective. Furthermore, the evaluations utilized a number of different control groups, recall periods, and follow-up times. Grouping such different comparisons together could have been misleading, but separating the evaluations into sub-comparisons rendered the groupings too small for an informative meta-analysis.

Data availability: Another major obstacle to meta-analyzing these results was the underutilization and underreporting of primary outcomes and difficulty entering some results into RevMan. Data unavailability also made it difficult to assess publication bias.

Statistical: There were several statistical reasons why a meta-analysis may have been misleading. Meta-analyzing primary studies with analyses that did not account for dropouts may have compounded this error, particularly since dropouts sometimes differed significantly from trial completers. Other statistical obstacles included accommodating cluster randomization, adjusting results for baseline differences, and handling trials with more than two arms, although these could have been overcome had we decided to proceed with the meta-analysis.

For these reasons, we present results from individual trials instead of providing a quantitative synthesis. RevMan analyses are provided where possible. Where possible, we transformed the results of cluster-randomized trials (Blake 2001, Clark 2005, Goldfarb 1999, Kirby 1997a, Kirby 1997b) before entering them into RevMan. Since Goldfarb 1999 did not report numbers of participants enrolled at baseline, average cluster size was calculated using the number of participants analyzed at follow-up; we calculated average cluster size for Kirby 1997a and Kirby 1997b using the available baseline enrollment numbers.

Analyses were impossible in RevMan for the following comparisons: Hernandez 1990; Miller 1993; continuous outcomes for Kirby 1997a, Kirby 1997b, and Kirby 1997c; and number of partners for Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, and Trenholm 2007d. These results are reported as available in the primary studies. All results are summarized in Figures 1-9 (see Figure 1).

Throughout these results, odds ratios less than one indicate a protective effect of the abstinence-only programs. Odds ratios are displayed with 95% confidence intervals. Results were significant at $p<0.05$.

Biological outcomes

Sexually transmitted infection

See Figure 3. Seven trials, with analyses representing 9779 participants at any time point, assessed participants' reports of ever having been diagnosed with a sexually transmitted infection by a doctor or nurse (Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm

2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). All compared an abstinence-only curriculum against usual care. No trial found a protective effect at any time point, and one trial found significant adverse effects compared to usual care.

Figure 3. 01. STI Incidence.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Kirby 1997a	Usual Care	Ever diagnosed with STD, among participants who reported never having been diagnosed at baseline	3	18%	1895 peer-led	Cluster-adjusted OR=3.04 (0.59, 15.73), P=0.18. Unadjusted for clustering, OR=2.43 (0.61, 9.75), P=0.21.
					2711 adult-led	**Cluster-adjusted OR=4.16 (1.16, 14.94), P=0.03. Unadjusted for clustering, OR=3.68 (1.20, 11.32), P=0.02.**
			17	34%	1545 peer-led	Cluster-adjusted OR=2.06 (0.67, 6.32), P=0.21. Unadjusted for clustering, OR=2.14 (0.77, 5.92), P=0.14.
					2313 adult-led	**Cluster-adjusted OR=2.73 (1.05, 7.14), P=0.04. Unadjusted for clustering, OR=2.85 (1.18, 6.89), P=0.02.**
Kirby 1997b	Usual Care	Ever diagnosed with STD, among participants who reported never having been diagnosed at baseline	17	26%	3761	Cluster-adjusted OR=0.77 (0.29, 2.09), P=0.61. Unadjusted for clustering, OR=0.86 (0.51, 1.47), P=0.59.
Kirby 1997c	Usual Care	Ever diagnosed with STD, among participants who reported never having been diagnosed at baseline	17	45%	372	OR=0.31 (0.03, 3.03), P=0.32.
Trenholm 2007a	Usual Care	Ever diagnosed with STD, among participants who reported ever having had sex	62.5	19%	277	OR=0.99 (0.28, 3.46), P=0.99
Trenholm 2007b	Usual Care	Ever diagnosed with STD, among participants who reported ever having had sex	65	20%	277	OR=1.46 (0.48, 4.49), P=0.50
Trenholm 2007c	Usual Care	Ever diagnosed with STD, among participants who reported ever having had sex	62.5	18%	163	OR=1.73 (0.35, 8.64), P=0.50
Trenholm 2007d	Usual Care	Ever diagnosed with STD, among participants who reported ever having had sex	59	16%	323	OR=0.83 (0.28, 2.42), P=0.73

Three trials (Kirby 1997a, Kirby 1997b, and Kirby 1997c) assessed STI diagnosis among participants who reported never having been diagnosed with an STI at baseline. Kirby 1997b and Kirby 1997c observed no significant findings at long-term follow-up (n=4133). Kirby 1997a systematically varied the staff delivering the intervention between adult and peer leaders; participants who received the intervention from adult leaders were significantly more likely than

controls to report having been diagnosed with an STI at three-month follow-up (n=2711, OR=4.16 [1.16, 14.94], p=0.03) and 17-month follow-up (n=2313, OR=2.73 [1.05, 7.14], p=0.04). It was unclear, however, whether this was due to changes in reporting, frequency of testing, or actual risk behavior. The youth-led arm of Kirby 1997a did not report significantly different rates of STI infection at three-month (n=1895, OR=3.04 [0.59, 15.73], p=0.18) or 17-month (n= 1545, OR=2.06 [0.67, 6.32], p=0.21) follow-up.

Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, and Trenholm 2007d assessed STI diagnosis among participants who reported ever having had sex. No trial found a significant effect at long-term follow-up compared to usual care (n=1040).

For the six studies with non-significant findings, odds ratios ranged from 0.31 (Kirby 1997c) to 1.73 (Trenholm 2007d). Long-term findings for Kirby 1997a and Kirby 1997c are limited by high attrition. Trial-specific results can be found in Figure 3.

Pregnancy

See Figure 4. Eight trials evaluated self-reported pregnancy incidence among 9417 participants (Anderson 1999, Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). No study found a protective at any time point, and one found a significant adverse effect compared to usual care.

Figure 4. 02. Pregnancy.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Anderson 1999	No treatment	Pregnancy since baseline among all participants	PI	0%	405	No events in either group, significance not estimable in RevMan
			12	38%	251	Unadjusted for clustering, OR=0.35 (0.02, 5.73), P=0.46. One event in each group.
Kirby 1997a	Usual Care	Ever pregnant or caused pregnancy, among participants who reported never having been pregnant or caused a pregnancy at baseline	17	34%	1548 youth-led	**Cluster-adjusted OR=2.35 (1.12, 4.95), P=0.02. Unadjusted for clustering, OR=2.49 (1.26, 4.90), P=0.008.**
					2295 adult-led	Cluster-adjusted OR=1.35 (0.73, 2.52), P=0.34. Unadjusted for clustering, OR=1.37 (0.78, 2.41), P=0.27.
Kirby 1997b	Usual Care	Ever pregnant or caused pregnancy, among participants who reported never having been pregnant or caused a pregnancy at baseline	17	26%	3758	Cluster-adjusted OR=1.36 (0.64, 2.90), P=0.43. Unadjusted for clustering, OR=1.36 (0.90, 2.03), P=0.14.
Kirby 1997c	Usual Care	Ever pregnant or caused pregnancy, among participants who reported never having been pregnant or caused a pregnancy at baseline	17	45%	371	OR=0.96 (0.13, 6.91), P=0.97.
Thenholm 2007a	Usual Care	Ever pregnant or caused pregnancy, among participants who reported ever having had sex	62.5	19%	277	OR=1.04 (0.37, 2.90), P=0.94
Thenholm 2007b	Usual Care	Ever pregnant or caused pregnancy, among participants who reported ever having had sex	65	20%	277	OR=0.97 (0.52, 1.78), P=0.91
Thenholm 2007c	Usual Care	Ever pregnant or caused pregnancy, among participants who reported ever having had sex	62.5	18%	163	OR=1.36 (0.41, 4.49), P=0.61
Thenholm 2007d	Usual Care	Ever pregnant or caused pregnancy, among participants who reported ever having had sex	59	16%	323	OR=1.33 (0.56, 3.16), P=0.52

Anderson 1999 assessed pregnancy since baseline among all participants, finding no significant effect compared to no treatment at short-term follow-up (n=405) or long-term follow-up (n=251). Three trials assessed pregnancy incidence ever, among participants who had reported never having been pregnant or caused a pregnancy at baseline (Kirby 1997a, Kirby 1997b, Kirby 1997c). Kirby 1997b and Kirby 1997c found no significant effects at long-term follow-up compared to usual care (n=4129). In Kirby 1997a, classrooms who received the intervention from peer leaders were significantly more likely than usual care controls to report having been pregnant at 17-month follow-up (n=1548, OR=2.35 [1.12, 4.95], p=0.02). However, this effect was largely influenced by the affirmative reports of six males from the seventh-grade class of a single school, and it was not mirrored by significant changes in any measured sexual behavior. There was no significant program effect for the adult-led sub-group of Kirby 1997a (n=2295, OR= 1.35 [0.73, 2.52], p=0.34).

Four trials assessed pregnancy incidence ever among participants who reported ever having had sex (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). No trial found a significant effect at long-term follow-up compared to usual care (n= 1040).

Odds ratios in studies reporting non-significant findings ranged from 0.35 (Anderson 1999 without controlling for clustering) to 1.36 (Kirby 1997b and Trenholm 2007c). Long-term findings for Anderson 1999, Kirby 1997a, and Kirby 1997c are limited by high attrition. Trial-specific findings appear in Figure 4.

Behavioral outcomes

Incidence of unprotected vaginal sex

See Figure 5. Five trials provided sufficient data to extract incidence of unprotected vaginal sex among all participants (Goldfarb 1999, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). Every trial compared an abstinence-only program to usual care. This outcome was assessed among all participants, including those who did not report recent sex; analyses represented 2892 participants at any time point. In analyses that controlled for clustering, no trial found a significant program effect at short-term or long-term follow-up.

Figure 5. 03. U Vaginal Sex.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Goldfarb 1999	Usual Care	Unprotected sex in past 1 month among all participants	2	unclear	839	Cluster-adjusted OR=0.56 (0.25, 1.23), P=0.15. Unadjusted for clustering, OR=0.54 (0.34, 0.84), P=0.007.
Trenholm 2007a	Usual Care	Unprotected sex in past 12 months among all participants	62.5	19%	447	OR=1.01 (0.67, 1.54), P=0.96
Trenholm 2007b	Usual Care	Unprotected sex in past 12 months among all participants	65	20%	479	OR=1.01 (0.67, 1.51), P=0.96
Trenholm 2007c	Usual Care	Unprotected sex in past 12 months among all participants	62.5	18%	413	OR=1.09 (0.60, 1.99), P=0.78
Trenholm 2007d	Usual Care	Unprotected sex in past 12 months among all participants	59	16%	714	OR=1.16 (0.76, 1.77), P=0.50

Goldfarb assessed unprotected sex among all participants in the past month, and results were limited to two-month follow-up. When we conducted analyses on the level of the individual, results significantly favored the program group over usual care (n=839, OR=0.54 [0.34, 0.84], p=0.007). However, when we re-analyzed these data using ICC=0.015 to control for cluster randomization, the effect was no longer significant (OR=0.56 [0.25, 1.23], p=

0.15).

We extracted outcome data from the remaining four studies by examining the percentage of participants who had sex in the last year and reported inconsistent condom use over the same period (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). No trial found a significant effect at long-term follow-up (n=2053). Odds ratios ranged from 1.01 (Trenholm 2007a) to 1.16 (Trenholm 2007d).

Trial-specific findings appear in Figure 5.

Incidence of vaginal sex

See Figure 6. Seven evaluations reported incidence of recent vaginal sex at immediate post-intervention follow-up, with analyses representing 3454 participants (Blake 2001, Clark 2005, Goldfarb 1999, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). Analyses represent all participants who were followed up in each trial. In analyses that controlled for clustering, one trial observed protective effects compared to usual care at short-term follow-up.

Figure 6. 04. Vaginal Sex.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Blake 2001	Non-Enhanced	Sex in past 3 months among all participants	1.5	10%	351	Cluster-adjusted OR=0.85 (0.24, 2.99), P=0.79. Unadjusted for clustering, OR=0.72 (0.24, 2.18), P=0.56.
Clark 2005	Usual Care	Sex since previous assessment among all participants	4.4	15%	211	Sex since baseline: Cluster-adjusted OR=0.69 (0.36, 1.32), P=0.26. Unadjusted for clustering, OR=0.68 (0.37, 1.23), P=0.20. Significant in paper using hierarchical level modeling ($p < 0.015$)
			12	37%	156	Sex in past 7.6 months: Cluster-adjusted OR=0.75 (0.37, 1.50), P=0.42. Unadjusted for clustering, OR=0.73 (0.38, 1.38), P=0.33.
Goldfarb 1999	Usual Care	Sex in past 1 month among all participants	2	unclear	839	**Cluster-adjusted OR=0.53 (0.29, 0.97), P=0.04. Unadjusted for clustering, OR=0.54, (0.38, 0.76), P=0.0004.**
Trenholm 2007a	Usual Care	Sex in past 12 months among all participants	62.5	19%	447	OR=0.95 (0.64, 1.40), P=0.79
Trenholm 2007b	Usual Care	Sex in past 12 months among all participants	65	20%	479	OR=0.81 (0.57, 1.17), P=0.27
Trenholm 2007c	Usual Care	Sex in past 12 months among all participants	62.5	18%	413	OR=1.10 (0.72, 1.70), P=0.66
Trenholm 2007d	Usual Care	Sex in past 12 months among all participants	59	16%	714	OR=1.09 (0.80, 1.50), P=0.57

Goldfarb 1999 evaluated sex in the past month among all participants; in analyses that controlled for clustering, findings significantly favored the abstinence-only program over usual care at two-month follow-up (n=839, OR=0.53 [0.29, 0.97], p=0.04). Measurement error was possible for this finding, because a higher percentage of control participants reported having sex in the past month than reported ever having sex.

Blake 2001 assessed sex in the past three months among all participants, and found no significant effect at 1.5-month follow-up compared to a non-enhanced version of the experimental program (i.e., without parent-child homework, n=351). Findings may not accurately evaluate program effect because the recall period encompassed both pre-intervention and post-intervention behaviors.

Clark 2005 assessed sex since baseline at 4.4-month follow-up (n=211), and sex since the 4.4-month follow-up at 12 months (n=156); findings were not significant in RevMan at either follow-up compared to usual care, regardless of whether we adjusted for clustering. However, using hierarchical level modeling, the paper reported a significant effect favoring the intervention group at short-term follow-up (p<0.015).

Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, and Trenholm 2007d assessed sex in the past 12 months among all participants at long-term follow-up, and none found a significant effect compared to usual care (n=2053).

Odds ratios in trials with non-significant findings ranged from 0.69 (Clark 2005) to 1.10 (Trenholm 2007c). Long-term results for Clark 2005 were limited by high attrition. Trial-specific results appear in Figure 6.

Frequency of vaginal sex

See Figure 7. Four studies assessed participants' frequency of vaginal sex, with analyses representing 2376 participants at any time point; no study observed protective effects at any time point, and one found an iatrogenic effect compared to usual care at short-term follow-up.

Figure 7. 05. Fq Vag Sex.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Hernandez 1990	Abstinence-Plus, Safer-Sex, No treatment	Number of times had sex in last 1.5 months among all participants	1.5	5%	388	NS in paper from repeated measures 4x2x2x2 factorial design (program x gender x sexual activity at baseline x time). Test statistics not available. Rank-ordered post-test means: Abstinence-only 175.6, Safe-Sex 192.9, Abstinence-plus 177.0, Nonintervention 197.2. Based on rank scores with range 1-366. Post-hoc comparisons not provided.
Kirby 1997a	Usual Care	Mean change in number of acts in last 3 months among sexually experienced at baseline or follow-up	3	18%	338 youth-led	**Means 0.9 vs. 0.3, P=0.02 from t-test based on change scores.**
					464 adult-led	Means 0.8 vs. 0.4, P=0.11 from t-test based on change scores.
			17	34%	338 youth-led	Means 1.8 vs. 1.6, P=0.62 from t-test based on change scores.
					586 adult-led	Means 2.0 vs. 1.6, P=0.16 from t-test based on change scores.
Kirby 1997b	Usual Care	Mean change in number of acts in last 3 months among sexually experienced at baseline or follow-up	17	26%	1012	Means 1.7 vs. 1.9, P=0.53 from t-test based on change scores.
Kirby 1997c	Usual Care	Mean change in number of acts in last 3 months among sexually experienced at baseline or follow-up	17	45%	52	Means 2.0 vs. 1.9, P=0.96 from t-test based on change scores.

Hernandez 1990 used a four-arm study to compare participants in abstinence-only, abstinence-plus, safer-sex, and non-intervention conditions. At six weeks after baseline (n=388), results indicated no significant program effects on the frequency of vaginal sex in the past six weeks according to a 4 x 2 x 2 x 2 factorial design (program assignment x gender x sexual activity at baseline x time). Kirby 1997a, Kirby 1997b, and Kirby 1997c examined the mean change in frequency of vaginal sex using a recall period of three months. Analyses were limited to participants who reported ever having had sex, and results are from t-tests based on change scores. Kirby 1997b and Kirby 1997c found no significant effects compared to usual care (n=1064). At three-month follow-up, the youth-led arm of Kirby 1997a reported a significantly larger increase in frequency of sex than usual care controls (n=338, means 0.9 vs. 0.3, p=0.02 in the primary study). The paper did not report this as a significant result due to a Bonferroni correction for multiple statistical tests. The effect was not significant at the 17-month assessment (n=338, means 1.8 vs. 1.6, p=0.62). No significant effects overall were found in the adult-led arm of Kirby 1997a at three-month follow-up (n=464, means 0.8 vs. 0.4, p=0.11) or 17-month follow-up (n=586, means 2.0 vs. 1.6, p=0.16).

Long-term findings for Kirby 1997a and Kirby 1997c are limited by high attrition. Results could not be analyzed in RevMan due to missing data. Trial-specific findings appear in Figure 7.

Number of sex partners

See Figure 8. Eight trials reported participants' number of sex partners, with analyses representing 4483 participants at any time point. No trial found a significant effect at any time point.

Figure 8. 06. Sex Partners.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Hernandez 1990	Abstinence-Plus, Safer-Sex, No treatment	Number of partners in last 1.5 months among all participants	1.5	5%	388	NS in paper from repeated measures 4x2x2x2 factorial design (program x gender, x sexual activity at baseline x time). Test statistics not available. Rank-ordered post-test means: Abstinence-only 180.8, Safe-Sex 190.6, Abstinence-plus 181.1, Nonintervention 181.3. Based on rank scores with range 1-366. Post-hoc comparisons not provided.
Kirby 1997a	Usual Care	Mean change in number of partners ever among sexually experienced at baseline or follow-up	3	18%	342 youth-led	Means 1.1 vs. 0.7, P=0.07 from t-test based on change scores.
					470 adult-led	Means 1.0 vs. 0.8, P=0.12 from t-test based on change scores.
			17	34%	393 youth-led	Means 2.3 vs. 2.0, P=0.28 from t-test based on change scores.
					584 adult-led	Means 1.9 vs. 1.8, P=0.64 from t-test based on change scores.
Kirby 1997b	Usual Care	Mean change in number of partners ever among sexually experienced at baseline or follow-up	17	26%	1012	Means 1.9 vs. 2.0, P=0.42 from t-test based on change scores.
Kirby 1997c	Usual Care	Mean change in number of partners ever among sexually experienced at baseline or follow-up	17	45%	53	Means 1.0 vs. 1.4, P=0.60 from t-test based on change scores.
Trenholm 2007a	Usual Care	Number of partners ever among all participants	62.5	19%	447	No significant effect, P=0.20 in paper from F-test of distributional difference
Trenholm 2007b	Usual Care	Number of partners ever among all participants	65	20%	479	No significant effect, P=0.80 in paper from F-test of distributional difference
Trenholm 2007c	Usual Care	Number of partners ever among all participants	62.5	18%	413	No significant effect, P=0.90 in paper from F-test of distributional difference
Trenholm 2007d	Usual Care	Number of partners ever among all participants	59	16%	714	No significant effect, P=0.49 in paper from F-test of distributional difference

Hernandez 1990 compared an abstinence-only program to an abstinence-plus program, a safer-sex program, and no treatment; the trial assessed participants' number of partners since baseline at six-week follow-up (n=388). The trial found no significant effect according to a 4 x 2 x 2 x 2 factorial design (program assignment x gender x sexual activity at baseline x time).

Three trials assessed participants' mean change in lifetime number of sex partners since baseline, with analyses limited to sexually experienced participants. These trials found no significant effect compared to usual care at short-term follow-up (n=812, Kirby 1997a), or at long-term follow-up (n=2042, Kirby 1997a, Kirby 1997b, Kirby 1997c).

Four trials assessed lifetime number of sex partners among all participants; according to F-tests of distributional differences, none found a significant effect at long-term follow-up compared to usual care (n=2053, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d).

Long-term findings for Kirby 1997a and Kirby 1997c are limited by high attrition. Results could not be analyzed in RevMan due to missing data. Trial-specific findings appear in Figure 8.

Lack of condom use

See Figure 9. We transformed condom use data to reflect lack of condom use, so that all odds ratios less than 1 would continue to indicate a protective effect. Eight trials assessed lack of condom use, with analyses representing 3254 participants at any time point; none found a significant program effect at short-term or long-term follow-up.

Figure 9. 07. No Condom Use.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Goldfarb 1999	Usual Care	Did not use a condom in past 1 month among participants reporting sex in the past month. (Unclear reporting: Could also have meant "did not use a condom ever" among participants reporting sex ever.)	2	unclear	171	Cluster-adjusted OR=0.98 (0.33, 2.91), P=0.97. Unadjusted for clustering, OR=0.88 (0.47, 1.62), P=0.67.
					167	If results referred to condom use ever among participants reporting sex ever, cluster-adjusted OR=0.75 (0.26, 2.23), P=0.61. Unadjusted OR=0.90 (0.49, 1.67), P=0.74.
Kirby 1997a	Usual Care	Did not use a condom at last act of sexual intercourse among participants who reported ever having had sex at baseline or follow-up	3	18%	339 youth-led	Cluster-adjusted OR=0.88 (0.55, 1.43), P=0.62. Unadjusted for clustering, OR=0.89 (0.57, 1.37), P=0.59
					471 adult-led	Cluster-adjusted OR=0.67 (0.44, 1.01), P=0.06. Unadjusted for clustering, OR=0.67 (0.46, 0.97), P=0.03
			17	34%	394 youth-led	Cluster-adjusted OR=1.18 (0.75, 1.84), P=0.48. Unadjusted for clustering, OR=1.19 (0.80, 1.79), P=0.39
					584 adult-led	Cluster-adjusted OR=0.95 (0.66, 1.39), P=0.80. Unadjusted for clustering, OR=0.96 (0.68, 1.34), P=0.80
Kirby 1997b	Usual Care	Did not use a condom at last act of sexual intercourse among participants who reported ever having had sex at baseline or follow-up	17	26%	1012	Cluster-adjusted OR=1.02 (0.62, 1.67), P=0.93. Unadjusted for clustering, OR=1.04 (0.80, 1.35), P=0.79
Kirby 1997c	Usual Care	Did not use a condom at last act of sexual intercourse among participants who reported ever having had sex at baseline or follow-up	17	45%	53	OR=1.22 (0.39, 3.79), P=0.73.
Trenholm 2007a	Usual Care	Did not use a condom at first vaginal sex among participants reporting ever having had sex	62.5	19%	277	OR=0.99 (0.50, 1.97), P=0.97
Trenholm 2007b	Usual Care	Did not use a condom at first vaginal sex among participants reporting ever having had sex	65	20%	277	OR=0.60 (0.26, 1.37), P=0.22
Trenholm 2007c	Usual Care	Did not use a condom at first vaginal sex among participants reporting ever having had sex	62.5	18%	163	OR=0.80 (0.31, 2.03), P=0.63
Trenholm 2007d	Usual Care	Did not use a condom at first vaginal sex among participants reporting ever having had sex	59	16%	323	OR=1.21 (0.69, 2.10), P=0.50

Goldfarb 1999 compared an abstinence-only program to usual care and reported the following at two-month follow-up: "For those reporting intercourse, 47% of controls and 50% of the intervention group reported using a condom." It was unclear whether this referred to condom use at last sex, condom use in the past month, or condom use ever. It was also unclear whether this referred to participants who had sex in the past month or participants who had ever had sex. We transformed this analysis to reflect "lack of condom use," and we analyzed this data first among participants who reported sex in the past month (n=171, OR=0.98 [0.33, 2.91], p=0.97), and then among participants who reported sex ever (n=167, OR=0.75 [0.26, 2.23], p=0.61). Neither finding was significant.

Three trials assessed lack of condom use at last sex among sexually experienced participants, and found no significant effect compared to usual care at short-term follow-up (n=810, Kirby 1997a) or long-term follow-up (n= 2043, Kirby 1997a, Kirby 1997b, Kirby 1997c). Results at three-month follow-up for the adult-led arm of Kirby 1997a initially favored the intervention group when we conducted analyses on an individual level (n=471, OR=0.67 [0.46, 0.97], p=0.03), but were not significant once we adjusted for clustering (OR=0.64 [0.44, 1.01], p=0.06). The paper does not report this finding as significant due to a Bonferroni correction.

Four trials found no significant effect on lack of condom use at first sex at long-term follow-up, among participants who reported ever having had sex (n=1040, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d).

Specific odds ratios for lack of condom use ranged from 0.60 (Trenholm 2007b) to 1.22 (Kirby 1997c). Long-term findings for Kirby 1997a and Kirby 1997c are less robust due to high attrition. Trial-specific findings appear in Figure 9.

Additionally, one trial assessed the absolute number of times participants used condoms in the past six weeks, and found no significant difference between participants in abstinence-only, abstinence-plus, safer-sex, and no-treatment conditions (n=388, Hernandez 1990). However, this outcome does not indicate the proportion of sex acts for which condoms were used; it simply reports the number of times participants used condoms at all.

Sexual initiation

See Figure 10. Ten studies reported the incidence of sexual initiation (i.e., whether participants had ever had sex), with analyses representing 11298 participants at any time point. No study found a significant effect at short-term or long-term follow-up.

Figure 10. 08. Initiation.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Blake 2001	Non-Enhanced	Ever had sex among all participants	1.5	10%	351	Cluster-adjusted OR=0.74 (0.28, 1.98), P=0.55. Unadjusted for clustering, OR=0.69 (0.29, 1.64), P=0.40.
Goldfarb 1999	Usual Care	Ever had sex among all participants	2	unclear	839	Cluster-adjusted OR=0.66 (0.36, 1.21), P=0.18. Unadjusted for clustering, OR=0.64 (0.45, 0.90), P=0.01.
Trenholm 2007a	Usual Care	Ever had sex among all participants	62.5	19%	447	OR=1.01 (0.68, 1.50), P=0.97
Trenholm 2007b	Usual Care	Ever had sex among all participants	65	20%	479	OR=0.85 (0.59, 1.22), P=0.38
Trenholm 2007c	Usual Care	Ever had sex among all participants	62.5	18%	413	OR=1.10 (0.72, 1.67), P=0.65
Trenholm 2007d	Usual Care	Ever had sex among all participants	59	16%	714	OR=1.17 (0.87, 1.58), P=0.29
Clark 2005	Usual Care	Initiation among participants who reported being sexually inexperienced at baseline	4.4	15%	134	Cluster-adjusted OR=0.28 (0.07, 1.10), P=0.07. Unadjusted for clustering, OR=0.25 (0.07, 0.97), P=0.04.
			12	37%	101	Cluster-adjusted OR=0.53 (0.17, 1.66), P=0.28. Unadjusted for clustering, OR=0.50 (0.18, 1.42), P=0.19.
Kirby 1997a	Usual Care	Initiation among participants who reported being sexually inexperienced at baseline	3	18%	1678 youth-led	Cluster-adjusted OR=1.10 (0.69, 1.75), P=0.70. Unadjusted for clustering, OR=1.10 (0.72, 1.67), P=0.67.
					2435 adult-led	Cluster-adjusted OR=1.04 (0.70, 1.52), P=0.86. Unadjusted for clustering, OR=1.03 (0.73, 1.46), P=0.87.
			17	34%	1431 youth-led	Cluster-adjusted OR=1.07 (0.78, 1.46), P=0.68. Unadjusted for clustering, OR=1.07 (0.80, 1.41), P=0.66.
					2134 adult-led	Cluster-adjusted OR=0.86 (0.67, 1.12), P=0.27. Unadjusted for clustering, OR=0.86 (0.68, 1.09), P=0.21.
Kirby 1997b	Usual Care	Initiation among participants who reported being sexually inexperienced at baseline	17	26%	3446	Cluster-adjusted OR=1.15 (0.82, 1.59), P=0.42. Unadjusted for clustering, OR=1.14 (0.95, 1.36), P=0.15.
Kirby 1997c	Usual Care	Initiation among participants who reported being sexually inexperienced at baseline	17	45%	362	OR=0.91 (0.42, 1.94), P=0.80.

Six trials assessed all participants' reports of ever having sex, and none found a significant effect at short-term follow-up compared to a non-enhanced program version (without parent-child homework, n=351, Blake 2001), at short-term follow-up compared to usual care (n=839, Goldfarb 1999), or at long-term follow-up compared to usual care (n=2053, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). Results for Goldfarb 1999 were initially significant without controls for clustering (n=839, OR=0.64 [0.45, 0.90], p=0.01), but findings did not remain significant with adjustments for clustering (OR=0.66 [0.36, 1.21], p=0.18).

Four trials assessed sexual initiation among participants who reported never having had sex at baseline; all four compared an abstinence-only program to usual care. None found a significant effect at short-term follow-up (n=4247, Clark 2005, Kirby 1997a), or at long-term follow-up (n=7474, Clark 2005, Kirby 1997a, Kirby 1997b, Kirby 1997c). When we analyzed results for Clark 2005 on an individual level, findings significantly favored the intervention group (n=134, OR=0.25 [0.07, 0.97], p=0.04), but were no longer significant once we adjusted for clustering (OR=0.28 [0.07, 1.10], p=0.07).

Odds ratios across trials ranged from 0.28 (Clark 2005) to 1.17 (Trenholm 2007d). Long-term findings for Clark 2005, Kirby 1997a, Kirby 1997c are limited by high attrition. Trial-specific findings appear in Figure 10.

Sexual risk behavior index

See Figure 11. Miller 1993 was the only study to use a risk behavior index, which included behaviors ranging from hand-holding to sexual intercourse or "the sexual act by which pregnancy can occur." The study was a three-arm trial: the three arms received (1) a video-based intervention complete with periodic newsletters, (2) the video-based intervention alone, or (3) no intervention. At 12-month follow-up, no significant group-by-time interaction was observed (p=0.662) according to a 3 (group) x 3 (time) repeated measures ANOVA. Reports of vaginal sex could not be disaggregated from this outcome, although the incidence of vaginal sex reportedly ranged from 3% to 5% in the three intervention arms, with a total of 503 participants at follow-up.

Figure 11. 09. Risk Index.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Miller 1993	Non-Enhanced, No treatment	Sexual behavior index with listed behaviors from hand-holding to vaginal sex	12	8%	503	NS in paper from group-by-time interaction for all 3 arms, P=0.662 according to 3 (group) x 3 (time) repeated measures ANOVA. Unclear if analyses controlled for baseline values.

Knowledge

See Figure 12. Seven trials reported any measure of knowledge, but none of these specifically assessed knowledge related to HIV/AIDS (Blake 2001, Goldfarb 1999, Miller 1993, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d).

Figure 12. 10. Knowledge.

Study	Comparison	Outcome Measured	Time (mo)	Attrition	N	Result
Blake 2001	Non-Enhanced	Perception of the effectiveness of abstinence as a preventive method	1.5	10%	351	Adjusted for clustering, OR=0.83 (0.42, 1.67), P=0.61. Unadjusted for clustering, OR=0.79 (0.43, 1.45), P=0.44.
		Knowledge of the risks of pregnancy the first time one has sexual intercourse	1.5	10%	351	Adjusted for clustering, OR=1.19 (0.48, 2.96), P=0.71. Unadjusted for clustering, OR=1.20 (0.54, 2.66), P=0.66.
Goldfarb 1999	Usual Care	"Sexuality knowledge": Unclear if this was HIV knowledge.	2	unclear	839	Intervention group showed significant increases in the number of correct answers to all four knowledge questions. the control group showed no significant change on these items.
Miller 1993	Non-Enhanced, No treatment	"Knowledge of sexual facts" (physical development, myths about teen sex, risks of sex, interpersonal respect"	12	8%	503	Intervention groups have "a significant increase in knowledge from pretest to posttest, with a plateau thereafter." Controls "showed a gradual increase in knowledge, with their delayed post-means having caught up with the treatment post-means."
Trenholm 2007a	Usual care	Identification of true and false STDs	62.5	19%	447	Means 83 vs. 73, p=0.00 from weighted regression model
		Knowledge of unprotected sex risks	62.5	19%	447	Means 0.98 vs. 0.94, P=0.04 from weighted regression model
		Knowledge of STD consequences	62.5	19%	447	Means 0.60 vs. 0.55, P=0.05 from weighted regression model
Trenholm 2007b	Usual care	Identification of true and false STDs	65	20%	479	Means 74 vs. 72, P=0.16 from weighted regression model
		Knowledge of unprotected sex risks	65	20%	479	Means 0.92 vs. 0.95, P=0.09 from weighted regression model
		Knowledge of STD consequences	65	20%	479	Means 0.56 vs. 0.56, P=0.90 from weighted regression model
Trenholm 2007c	Usual care	Identification of true and false STDs	62.5	18%	413	Means 63 vs. 65, P=0.45 from weighted regression model
		Knowledge of unprotected sex risks	62.5	18%	413	Means 0.88 vs. 0.86, P=0.47 from weighted regression model
		Knowledge of STD consequences	62.5	18%	413	Means 0.52 vs. 0.47, P=0.08 from weighted regression model
Trenholm 2007d	Usual care	Identification of true and false STDs	59	16%	714	Means 57 vs. 56, P=0.55 from weighted regression model
		Knowledge of unprotected sex risks	59	16%	714	Means 0.74 vs. 0.75, P=0.64 from weighted regression model
		Knowledge of STD consequences	59	16%	714	Means 0.40 vs. 0.44, P=0.07 from weighted regression model

Blake, et al assessed whether participants knew "the effectiveness of abstinence" and the risk of pregnancy at first intercourse (Blake 2001), and found no significant difference compared to a non-enhanced program version (without parent-child homework) at immediate post-intervention (n=351).

Goldfarb, et al assessed "sex knowledge" with a usual care control group at two-month follow-up (n=839), but did not report a measure of between-group significance (Goldfarb 1999). As the study reports, "the intervention group showed significant increases in the number of correct answers to all four knowledge questions ... the control group showed no significant change."

Miller assessed knowledge of sexual facts, including "physical development, myths about teen sex, risks of sex, [and] interpersonal respect" (Miller 1993). The trial compared an abstinence-only program with newsletters, a program without newsletters, and a no-treatment control, with assessments at immediate post-intervention and 12 months. Again, between-groups tests of significance were not provided. As the trial reports, intervention groups had "a significant increase in knowledge from pre-test to post-test, with a plateau thereafter," while controls "showed a gradual increase in knowledge, with their delayed post-means having caught up with the treatment post-means."

The four trials by Trenholm, et al compared different abstinence-only programs to usual care. These trials assessed three outcomes: identification of "true and false STDs," knowledge of the risks of unprotected sex, and knowledge of the consequences of STDs (Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). One trial found significantly protective effects for all three outcomes at 62.5-month follow-up (n=447; Trenholm 2001); the remaining three trials found no protective effects at long-term follow-up (n=1606; Trenholm 2007b, Trenholm 2007c, Trenholm 2007d) Additionally, an analysis of all four trials together suggested that the programs may have led to a significant adverse effect on knowledge of condoms. At long-term follow-up, the overall analysis showed that a larger proportion of intervention youth reported that condoms "never prevent HIV" compared to usual care controls (21% vs 17%); intervention youth were also significantly less likely than controls to report that condoms "usually" prevent HIV (34% vs. 38%).

Cost data and participant satisfaction

None of the evaluations reported cost-effectiveness, although PSI/ENABL, as evaluated in Kirby 1997a, Kirby 1997b, and Kirby 1997c was reportedly "expensive" (Saunders 1996, cited in Cagampang 1997). Overall, the PSI/ENABL program received $15 million for the first three-year funding cycle, including the cost of evaluation (Cagampang 1997). The three-part evaluation was presented to state administrators, including the governor, before the removal of state funding from the PSI program (Cagampang 1997).

Four studies collected data on participant satisfaction; Anderson 1999 reported that participant surveys, ethnographic observations, interviews, and focus groups indicated high levels of satisfaction and youth participant engagement. Kirby 1997a, Kirby 1997b, and Kirby 1997c collected participant satisfaction data from focus groups, interviews, and surveys, and similarly reported high levels of satisfaction (along with political popularity). Of the participants surveyed, 82% reported that the program was "good or excellent," while 88% stated that the program information was helpful. However, participants who had a "serious boyfriend or girlfriend" reported significantly lower levels of satisfaction than youth who had not had a serious relationship; additionally, 82% of participants who suggested additional program topics recommended that the program contain more information about "how to keep from getting STDs and HIV."

DISCUSSION

Summary of main results

As this review has highlighted, methodologically rigorous assessments of abstinence-only interventions for HIV prevention are hindered by a number of obstacles. Foremost among these are the scarcity of randomized controlled designs, the methodological deficiencies of included trials, the relatively homogeneous nature of participant groups, the underutilization of outcomes relevant to HIV risk, and incomplete reporting by primary trials. Even with these caveats, however, the evidence from this review has critical implications for research and practice.

Behavioral outcomes

The data from this review show no indications that abstinence-only programs can effectively reduce HIV risk as indicated by self-reported sexual behavior. Evidence from these primary trials suggests that when compared to a variety of control groups, the evaluated programs did not protectively or adversely affect incidence of unprotected vaginal sex, the frequency of vaginal sex, participants' number of sex partners, condom use, rates of sexual initiation, or overall sexual behavior according to a risk index.

Notably, findings do not suggest that abstinence-only programs can effectively encourage abstinent behavior; although programs did not appear to cause harm, the bulk of the evidence suggests that the programs are ineffective for preventing or decreasing sexual activity. This was true for both primary abstinence (i.e., preventing sexual initiation) and secondary abstinence (i.e., decreasing the incidence and frequency of recent sex). Although an isolated protective effect was observed for incidence of vaginal sex (n=839, Goldfarb 1999) it was offset by possible reporting error, a

lack of long-term follow-up data, and non-significant findings in six other trials among 2615 participants (Blake 2001, Clark 2005, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d). A significant adverse effect for frequency of recent sex (n= 338, Kirby 1997a) was likewise limited to short-term follow-up and offset by non-significant findings in other trials (see "Harms" below).

The evidence of behavioral effects is limited by several factors: quantitative synthesis of the available results was impossible; reporting bias or differing definitions of behavioral terms may impede reliable self-reports of behavior; behavioral outcomes are somewhat limited in their ability to indicate HIV risk (Peterman 2000, O'Leary 1997); floor effects may constrain the possibility of finding significant behavior change in participant populations with low levels of sexual activity (e.g. Miller 1993); three of the evaluations reporting behavioral outcomes were part of the same three-study design; three of the studies reporting behavioral outcomes had overall attrition rates exceeding 33%; and attrition was unavailable for one study (Goldfarb 1999).

Biological outcomes

There appears to be no evidence that abstinence-only programs can effectively reduce HIV risk as indicated by self-reported STI and pregnancy incidence. Eight trials offer some evidence to suggest that abstinence-only programs are ineffective in reducing these events on both a short-term and long-term basis, and Kirby 1997a provided some evidence of adverse program effects for both outcomes; however, these were offset by non-significant findings in the remaining seven trials (see "Harms" below).

The evidence of biological effects faces many of the same limitations: floor effects constrain the possibility of finding significant effects where prevalence is low; quantitative synthesis of the available results was impossible; reporting bias or the lack of diagnosis may impede reliable self-reports of STI and pregnancy (i.e., relying on diagnosis does not capture undiagnosed conditions, especially given that many STIs are asymptomatic); STI incidence and especially pregnancy incidence are somewhat limited in their ability to indicate HIV risk; seven of the eight evaluations reporting either of these outcomes were conducted by just two groups of trialists; and three of the studies reporting biological outcomes had overall attrition rates exceeding 33%.

Knowledge

Because no trial in this review specifically assessed the knowledge of HIV/AIDS facts, there is insufficient information from this review to gauge how abstinence-only programs may affect HIV/AIDS knowledge.

Harms

Several iatrogenic effects were observed throughout the review on STI incidence, pregnancy incidence, and the frequency of recent vaginal sex, all of which were documented in Kirby 1997a. As clarified in Figure 3, Figure 4, and Figure 7, each of these harms is offset by studies showing evidence of no effect, and long-term harms for STI incidence and pregnancy were offset by high attrition. Because the significant harms do not appear consistently, we conclude that abstinence-only programs do not appear to reliably cause harm on the behavioral and biological outcomes of interest. This review does not address psychological or attitudinal harms, which merit attention in future investigations; such studies might specifically investigate program effects on participants with baseline sexual experience, participants who have been subject to coerced sex, or participants experiencing same-sex attraction.

Specific behavioral harms observed in this review were as follows:

"Participants in the adult-led intervention arm were significantly more likely than usual care controls to report having ever been diagnosed with an STI at three-month follow-up (n=2711, OR= 4.16 [1.16, 14.94], p=0.03) and 17-month follow-up (n=2313, OR=2.73 [1.05, 7.14], p=0.04). These effects did not correspond to increases in self-reported sexual behavior, and it is unclear whether this resulted from differences in rates of testing, diagnosis, or actual STI incidence. This was offset by non-significant findings in the youth-led arm of Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, and Trenholm 2007d (total n=7068).

"Participants in the peer-led intervention arm were significantly more likely than usual care controls to report having been pregnant or caused a pregnancy at long-term follow-up (OR=2.47 [1.18, 5.17]). This outcome did not mirror any corresponding long-term change in self-reported sexual behavior, and further analyses revealed that effects could be tracked to male participants in the same class of a single school, suggesting possible reporting inaccuracies or other alternative explanations (see Kirby 1997a). This was offset by non-significant findings in Anderson 1999, the adult-led arm of Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, and Trenholm 2007d (total n=7869).

"At short-term follow-up, sexually experienced participants in the peer-led arm reported a significantly greater increase in the number of sex acts in the last three months than their usual care counterparts. This effect was not sustained at long-term follow-up, and it was offset by non-significant findings in Hernandez 1990, the adult-led arm of Kirby 1997a, Kirby 1997b, and Kirby 1997c (total n=2038).

Overall summary

In sum, the 13 trials reviewed here show no evidence that abstinence-only programs effectively decrease or exacerbate HIV risk among participants in high-income countries, as measured by self-reported biological and behavioral outcomes. There is some evidence to suggest that the programs are ineffective for improving

biological and behavioral outcomes, but the generalizability of this evidence may be limited to the US youth populations represented in the included trials.

Additional findings of ongoing and recently completed trials

Beyond the included results, our search discovered mid-term findings of two ongoing trials (Miller-Heyl 2007, Nagel 2007) and long-term findings of a recently completed trial (Jemmott 2006), but full reports were not yet available. No results were available for another two ongoing trials (Borawski 2007, Markham 2007). All trials enrolled US adolescents and are classified as "ongoing" in Figure 1 and the Table of Ongoing Studies. Based on preliminary results, we do not believe that including full reports of the three trials with follow-up data would have changed the conclusions of our review. At the time of this review, there was insufficient information available to assess the methodological quality of these ongoing studies, so this must be kept in mind when comparing their results to the findings of this review.

One ongoing cluster-randomized trial (Nagel 2007) is based in Toledo, Ohio, and randomized 510 adolescents to an eight-session school-based abstinence-only program or no treatment. Chi-square analyses at immediate post-test found no significant difference in participants' reports of vaginal sex in the past two months. No other behavioral outcomes were presented. These findings reinforce the non-significant results of six included trials from our review (n=2,615), compared to one included trial with significant findings (n=839).

A multi-site ongoing trial (Miller-Heyl 2007) randomized the families of 189 adolescents in Denver and Montezuma County, Colorado, to a 22-hour community-based abstinence-only program or no treatment. At six-month follow-up (n=132), there was no significant difference in whether participants had ever had sex (p= 0.15 from a group-by-time interaction, using a repeated-measures ANOVA). This aligns with the non-significant findings of ten included trials from our review (n=11,298). Results of the ongoing trial at 12-month follow-up favored the intervention group for this outcome, but significance was not stated.

The recently completed trial (Jemmott 2006) allocated 662 adolescents to ten trial arms spanning four conditions: abstinence-only (two arms), abstinence-plus (four arms), safer sex (two arms), and attention control (two arms). At 24-month follow-up (n= 559), logistic regression found that abstinence-only program participants were less likely to report ever having had sex than participants in the attention control (p=0.02), abstinence-plus (p=0.05), and safer-sex (p=0.007) conditions. With analyses limited to participants who reported never having had sex at baseline (n<559), effects remained significant compared to the attention control (p= 0.01) and the safer-sex program (p=0.007), but not compared to the abstinence-plus program (p=0.07). These findings are offset by the non-significant results of ten included trials in this review (n= 11,298). The trial also found no significant differences between the abstinence-only program and the attention control in consistent condom use or condom use at last sex at 24-month follow-up (n<224, p-value not reported), which aligns with non-significant findings in eight trials in this review (n=3,254). Condom use comparisons to the abstinence-plus and safer-sex arms were not reported and could not be obtained; however, previous trials evaluating variants of the abstinence-plus program have found significantly protective effects for condom use and unprotected sex compared to attention controls (Jemmott 1992, Jemmott 1998, Jemmott 1999, Jemmott 2004).

Quality of the evidence: internal validity and methodological rigor

Overall, the methodological quality of the 13 included studies was difficult to judge due to incomplete reporting of key methodological and clinical features. Reporting deficiencies may obscure important methodological strengths or weaknesses, and it was often difficult to obtain additional information by contacting the trialists.

As a group, the studies had several important strengths; these included controlling for baseline assessment values, describing the development or piloting of data collection tools, enrolling relatively large samples at baseline (median baseline enrolment=626), acknowledging the need for long-term follow-up data, and making efforts to improve the validity of self-reported data. The trial results for outcomes of interest were also remarkably consistent across studies. However, these strengths were partially overshadowed by the following methodological weaknesses or ambiguities:

"Underreporting of key methodological features: No evaluation specified procedures for allocation concealment, and few reported method of randomization. Randomization procedures may be questioned for the two studies that had unequal distributions of participants among trial arms, but did not specify an intention to do so (Anderson 1999, Miller 1993); one published trial report did not state clearly that randomization was used (Goldfarb 1999), although this was discovered in correspondence with the trialist. No study reported procedures for blinding the personnel who administered outcome surveys.

"Missing data: Commonly missing values across studies included the number of participants per trial arm at baseline and follow-up, means and standard deviations for continuous outcomes, percentages for dichotomous outcomes, effect sizes, and attrition analyses. Several trials were also missing statements of baseline equivalence (Miller 1993), total attrition and baseline enrolment (Goldfarb 1999), and data for the estimation of average cluster size (Anderson 1999). Only five trials reported statistical power in relation to sample size (Clark 2005, Trenholm 2007a, Trenholm 2007b, Trenholm 2007c, Trenholm 2007d); however, median sample size was fairly large, and we do not perceive this to be a major limitation.

"Attrition at last follow-up: Dropout exceeded 33% in four of the included studies (Anderson 1999, Clark 2005, Kirby 1997a,

Kirby 1997c) and was not reported in one study (Goldfarb 1999). No trial used an intention-to-treat analysis.

"Unit of analysis problems: Of the six trials that used cluster randomization, two studies did not report controlling for clustering in analyses (Anderson 1999, Goldfarb 1999), potentially leading to Type I errors. Our inability to estimate a reliable ICC from available study data is a possible source of bias for the review as a whole.

"Limitations of outcome measures: All outcome data in this review are vulnerable to self-report bias. While biological outcomes are the best indicators of HIV risk, there are obstacles to using these outcomes among youth populations. These include resistance from schools or parents, logistical difficulties, and floor effects (i.e., a recent study suggests that overall HIV prevalence in young adults in the US is 1.0 per 1000, Morris 2006). For these reasons, self-reported behavioral outcomes were a necessary proxy for HIV risk. As discussed elsewhere, there are many limitations to self-reported sexual behavior data, which is an inevitable source of bias. The results of this review highlight the need for a standardized set of outcome measures with unambiguous and explicit definitions, consistent follow-up times and recall periods, and clinically meaningful implications for HIV risk. Reaching a consensus on a set of behavioral outcomes, follow-up times, recall periods, sub-groups of interest, and reporting formats would transform researchers' ability to make comparisons across primary trials. As this review demonstrates, behavioral outcomes of HIV prevention trials are often non-specifically defined (particularly for terms known to have varying meanings); different studies use differing follow-up times and recall periods; the most clinically meaningful outcomes (e.g. STI incidence, incidence of unprotected vaginal sex) are rarely used; outcomes are rarely separated into oral, anal, and vaginal sex, which carry different levels of HIV risk; outcomes are reported for various subsets of participants across studies; and long-term follow-up data are sometimes not collected. Assessing the incidence and frequency of unprotected oral and anal sex acts may be particularly relevant for participants of abstinence-only programs; recent research on virginity pledge programs (Bruckner 2005) suggests that pledgers may delay vaginal sex, but do not differ significantly from non-pledgers in rates of STI due to oral and anal sex. Long-term follow-up data are also particularly relevant for studies of youth sexual behavior, which is often characterized by intermittent periods of activity (Rotheram-Borus 2000).

"Insufficient reporting on program design and implementation: Incomplete reporting on the design and implementation of both experimental and control interventions has been previously noted as a limitation of behavioral HIV prevention trials (O'Leary 1997); ideally, reports of implementation should be sufficiently detailed to permit replication, or at least to permit reviewers to judge whether programs are similar enough to combine their evaluations in a systematic review (Montgomery 2005). Additionally, detailed reporting on implementation characteristics can allow the tracking of particular program components across studies, leading to the identification of potentially effective program characteristics. Few of the trials in this review reported this level of detail on intervention design, program delivery by staff, uptake by participants, and unique features of the trial context. Kirby 1997a, Kirby 1997b, Kirby 1997c, Trenholm 2007b, and Trenholm 2007c were exceptions to this trend; however, information on program design and implementation for these studies was contained in unpublished reports. Journals may wish to consider dedicating more space to the reporting of key implementation data for studies of behavioral HIV prevention programs. Details on the treatment of control groups were rarely reported, although these are crucial to understand trial comparisons. To address the current debate over abstinence-only programs, comparisons of abstinence-only programs against abstinence-plus, condom promotion, or safer-sex programs may provide the most relevant data. It is possible that studies with "usual care" control groups may have made some of these comparisons, but unclear reporting of what control group participants actually received made this difficult to determine.

Overall completeness and applicability of evidence

External validity

The objective of this review was to determine the effectiveness of abstinence-only programs for HIV prevention among all possible participants in high-income countries. The evidence gathered by our search is somewhat able to address this objective, but several factors limit the generalizability (external validity) of these results.

First, relatively little data were available for the primary outcomes of interest. Data were also insufficient to inform assessments of cost-effectiveness, participant satisfaction, implementation fidelity, intervention mechanisms, and the effectiveness of specific program characteristics across studies. The included program designs were consistent with our background research, but they may not represent the entire realm of abstinence-only approaches in high-income countries.

Second, the population represented in this review is limited to adolescents and young adults in the US. These participants may not be representative of the universe of high-income country populations; no evaluations were discovered in high-income countries other than the US. Even within the realm of US adolescents and young adults, a number of vulnerable groups are not represented in these trials. For example, data were unavailable or could not be disaggregated for gay, lesbian, bisexual, or transgender youth; youth with disabilities; youth who have immigrated to the US; substance-dependent youth; or homeless young people. Sub-group analyses by gender, ethnicity, age, socioeconomic status, family structure, religion, baseline sexual experience, or other variables were impossible. No data were available for participants younger than 9 or older than 21 years, and no study focused exclusively on youth in high school (aged 14-18 yrs). Several regions of the US were underrepresented, particularly the South and Midwest.

Third, one of the most salient research deficits in this review is the lack of direct comparisons of abstinence-only against abstinence-plus, condom-promotion, or safer-sex programs. According to our criteria for abstinence-only interventions, only Hernandez 1990 provided this type of comparison. Beyond these results and the findings of one ongoing study (Jemmott 2006), this review cannot directly address the differential effectiveness of abstinence-only and abstinence-plus strategies for HIV prevention. We will discuss this issue further in our Cochrane review of abstinence-plus programs for HIV prevention in high-income countries.

An additional study has compared an abstinence-focused program, a condom-focused program, and an attention control, suggesting that the condom-focused program had longer-lasting effects and more promising results among sexually experienced participants (Jemmott 1998). However, the abstinence-focused program in this evaluation acknowledged that condoms were an effective prevention strategy, which caused us to classify the program as an abstinence-plus intervention. This study has been incorporated into our review of abstinence-plus programs.

Evidence in the practice context

As mentioned in our Background section, examining the evidence from high-income countries is illuminating for several reasons. As compared to their counterparts in resource-poor countries, abstinence-only program participants in high-income countries have relatively higher levels of gender equality, lower levels of discrimination and structural violence, and greater access to education, employment, and healthcare resources. It is also relevant for this review that abstinence-only programs receive a great deal of federal, state, and private funding in the United States. We hypothesized that these conditions may offer abstinence-only programs in the US and other high-income countries the best possible chances of affecting risk behavior. A 2004 editorial in the Lancet concurred, suggesting that "abstinence only works where women have the means to make it work - an unobtainable position for many, if not most women living in some of the poorest regions of the world" (Lancet 2004). However, the results of this review suggest that even with relative material and social advantages, participants of the abstinence-only programs examined here did not alter their risk behavior as a result of the interventions.

Examining the evidence from the US might also allow a more informed assessment of current US policy with regard to abstinence-only interventions. In light of these results, it could be productive for policymakers and practitioners to reconsider current resource allocation. Although abstinence-only programs do not appear to consistently cause harm among participants in high-income countries, program evaluations also do not appear to show short-term or long-term reductions in risk behavior. This lack of protective effect may be cause to question the costs and benefits of abstinence-only approaches. To give some idea of abstinence-only funding levels, the 2007 budget provided by the White House allocates $204 million to domestic abstinence-only approaches, with the goal of increasing annual funding to $270 million by 2009 (OMB 2006). This estimate does not include state funding ($3 for every $4 in federal contributions), private contributions, or the funding for abstinence-only interventions in 15 at-risk countries via the President's Emergency Plan for AIDS Relief. PEPFAR regulations require 33% of the $3 billion allocated to prevention funding to be spent on abstinence-only interventions during the five-year plan (US 2003).

Should the funding of abstinence-only interventions continue at its current levels, policymakers and practitioners might consider allocating more of these resources to methodologically rigorous program evaluations with behavioral and biological outcomes. It is striking that, despite an exhaustive and highly sensitive search strategy, this review discovered randomized controlled evaluations of only 11 separate programs. Considering the political popularity, funding, and widespread implementation of abstinence-only programs, especially in the US (i.e., a recent three-wave survey of 15,170 young adults in the US found that 20% reported having taken a virginity pledge at one or more survey waves (Bruckner 2005)), we were surprised by the scarcity of rigorous evaluations with behavioral outcomes. We were especially concerned by the complete absence of evaluations of programs implemented by faith-based organizations, since these are among the most widespread channels for abstinence-only program delivery in the United States and elsewhere. Behavioral data from ongoing studies will help remedy some knowledge gaps, but the variety in abstinence-only program designs (see Wilson 2005) suggests that additional studies may be necessary for a clear and nuanced understanding of program effectiveness.

Potential biases in the review process

Although we have mentioned the limitations of our review throughout the previous sections, they are consolidated here for clarity. First, despite our extensive search for unpublished and ongoing trials (which included searching grey literature, handsearching conference abstracts, and contacting experts), our review remains vulnerable to publication bias. This review does not include studies indexed after February 2007. Our search for unpublished literature was somewhat hampered by non-response and the political connotations of terms such as "abstinence-only" and "abstinence-based;" it is possible that additional unpublished evaluations of abstinence-only programs exist, but were not called to our attention due to political concerns or varying definitions of these terms. Additionally, our electronic search utilized an HIV filter, which means that it did not reliably uncover evaluations of abstinence-only programs focused exclusively on pregnancy prevention. Several randomized controlled evaluations of abstinence-only pregnancy prevention programs were recovered by cross-referencing, but we classified these as excluded studies after determining that HIV prevention was not a stated goal. We suggest that users consult the pregnancy prevention literature for further

evidence of the effectiveness of abstinence-only programs for preventing pregnancy.

While assessing trials for inclusion, it was frequently difficult or impossible to assess the extent to which interventions emphasized abstinence, or how they discussed condom use or contraception. It is possible that we have omitted abstinence-only interventions that did not include any terms such as "postpone sex" or "abstain" in their titles or abstracts; key words for these evaluations may have been broad terms such as "sex education" or "family life education." Including terms with this broad scope in our search strategy would have made the review unfeasible due to the large number of search hits, so we have accepted this as a limitation of the review. We welcome suggestions to correct any errors of omission in future updates.

Beyond the limitations of our search strategy, we were often unable to obtain relevant missing data, including methodological characteristics, clinical characteristics, and outcome data. This led to several systematic gaps in the review, such as the inability to enter some comparisons into RevMan or to assess publication bias.

We also had difficulty finding a reliable intra-class correlation coefficient when computing the design effect for cluster-randomized trials. Although analyses investigated a range of intraclass correlation coefficients (ICCs), we report analyses according to previously published ICCs. To limit potential bias, we ultimately reported two sets of results for cluster-randomized trials in our charts: one set unadjusted for cluster randomization, and one set using the ICCs recommended for school-based studies in Johnson 2002. This affected the significance of results for Clark 2005 (sexual initiation at short-term follow-up), Kirby 1997a (lack of condom use at last sex at short-term follow-up) and Goldfarb 1999 (sexual initiation and recent vaginal sex). All differences were in the direction of non-significance in the cluster-adjusted findings.

Perhaps due to differences in analysis strategies or software, our methods of controlling for clustering, or our lack of original data sets, our re-analyzed results differed slightly from results published in several trials. These included Clark 2005 (incidence of vaginal sex at 4.4-month follow-up), Kirby 1997a (condom use at last sex at three-month follow-up for the adult-led arm), and Goldfarb 1999 (incidence of vaginal sex, sexual initiation). All differences were in the direction of non-significance in our re-analyzed results.

We did not use a Bonferroni or other correction to control for the examination of multiple statistical tests. In total, the results section of this review reports 79 discrete statistical tests (not including additional tests for cluster-randomized trials). Five of these tests (6%) reached significance at a $p<0.05$ level, which is what might be expected to achieve significance by chance.

Finally, we must acknowledge the potential for bias due to dominant cultural norms: two reviewers (KU and DO) are originally from the United States, while the third (PM) is from the United Kingdom.

Agreements and disagreements with other studies or reviews

Previous reviews have reached varying conclusions regarding the effectiveness of abstinence-only programs for the reduction of sexual risk behavior. Our review adds to previous assessments by virtue of its international focus; pre-specified, systematic, and highly sensitive search for program evidence; inclusion of published and unpublished literature; exclusive focus on behavioral and biological outcomes; pre-approved Cochrane protocol; and acceptance of only the most rigorous trial evidence (i.e., data from randomized and quasi-randomized controlled trials).

There is no evidence of abstinence-only program effectiveness in low- or middle- income countries.

First, our review found very similar results to the 2004 and 2006 reviews of abstinence-only programs for the prevention of HIV in developing countries. This review found only one abstinence-only program evaluation and "little or no evidence of effect on actual behavior" (O'Reilly 2004, O'Reilly 2006).

There is no evidence of abstinence-only program effectiveness in high-income countries.

Next, our review concurs with reviews carried out by NHS 1997, DiCenso 2002, Franklin 1997, Kirby 1997R, Kirby 2001, Kirby 2006, Manlove 2004, and Thomas 2000. Each of these reviews found no conclusive evidence from rigorous trials to suggest that abstinence-only interventions affect sexual risk behavior. We also agree with the assessment in DiCenso 2002 of the evidence from Kirby 1997a, Kirby 1997b, and Kirby 1997c, suggesting increased pregnancies in the partners of male participants.

Bennett 2005 included evaluations of three abstinence-only programs. Results suggested that one of the programs (Jorgensen 1993) had a significantly protective effect on sexual initiation; however, this program focused exclusively on pregnancy prevention. The two other abstinence-only evaluations in Bennett 2005 were Blake 2001 and the study encompassing Kirby 1997a, Kirby 1997b, and Kirby 1997c.

Grunseit 1997 included three abstinence-only trials: Christopher 1990, Jorgensen 1993, and Miller 1993. This review highlighted the iatrogenic effects observed in Christopher 1990, but did not provide a summary of overall effects of abstinence-only programs.

This review disagrees with Rector 2002, a non-systematic review of abstinence-only programs for US youth that concluded, "real abstinence programs can be highly effective in reducing early sexual activity." The methodological limitations of this review are summarized in Kirby 2002. Likewise, our review disagrees with the conclusions of Napier 1997, another non-systematic review of abstinence-only curricula for pregnancy prevention.

Abstinence-only programs are rare in high-income countries outside the United States.

Like a number of previous reviews, this review is based on findings of trials that enrolled US youth. However, this is the first review to systematically search for methodologically rigorous evidence from all high-income countries. The fact that we did not find methodologically rigorous program trials outside the US might indicate that such evaluations are inaccessible by existing search methods, or that abstinence-only programs are not popular HIV prevention strategies in other high-income countries. The second possibility appears likely, given the sensitivity of our search and several previous reviews suggesting that abstinence-based approaches are rare outside the US (Berne 1999, Jones 1985).

In general, HIV risk-reduction interventions have low rates of behavior change for abstinent behavior.

We concur with previous reviews suggesting relatively low rates of behavior change for abstinent behavior and the delay of sexual initiation as a result of HIV prevention interventions among adolescents (Jemmott 2000, Kim 1997, Pedlow 2003, Robin 2004). Our review differs slightly from Silva 2002, a meta-analysis of school-based programs for the promotion of abstinent behavior. This study included randomized and quasi-experimental trials of abstinence-only, abstinence-plus, and other school-based programs; it found a statistically significant effect for abstinent behavior, but the overall effect size was d=0.05 (95% confidence interval 0.01, 0.09) (Silva 2002). In sub-group analyses, Silva 2002 found no significant differences between "abstinence-oriented" programs (including abstinence-only and abstinence-plus designs) or safer-sex programs.

In general, HIV risk-reduction interventions do not significantly increase sexual risk behavior.

Finally, our review also concurs with a recent review and meta-analysis by Smoak, et al, which found that interventions targeting sexual risk behavior do not lead to systematic increases in sexual activity or number of partners (Smoak 2006). This review summarized 174 trials of diverse HIV risk reduction interventions; it was not limited to abstinence-only interventions or to high-income countries.

AUTHORS' CONCLUSIONS

Implications for practice

"Given the methodological limitations of the evidence, the homogeneity of the trial populations, the underutilization of primary outcomes, and the potential variety of abstinence-only strategies, this review cannot draw unequivocal conclusions about all abstinence-only programs in all high-income countries. We also cannot comment on program effectiveness in participant populations outside adolescents and young adults in the United States.

"With these caveats, the available studies show no evidence that abstinence-only programs are effective or harmful for the prevention of HIV among US youth, as measured by self-reported behavioral and biological outcomes. The 13 included trials suggested that abstinence-only programs are ineffective for reducing HIV risk, but the generalizability of these data may be limited.

"Current resource allocation to abstinence-only interventions in the US may not reflect the limited evidence for program effectiveness. Should funding for abstinence-only approaches continue at its current levels, a larger proportion of these resources might be productively invested in rigorous program evaluations with behavioral and biological outcomes.

Implications for research

Future research could address the following questions raised by this review:

"Direct comparisons between abstinence-only programs and other program types (e.g., abstinence-plus, condom promotion, safer-sex)

"Intervention mechanisms

"Potential attitudinal and psychological harms, especially among participants with sexual experience, participants who have experienced coerced sex, or participants who identify as gay, lesbian, bisexual, or transgendered

"Program effectiveness among especially vulnerable groups (e.g., gay, lesbian, bisexual, or transgendered youth; youth with disabilities; recent immigrants; substance-dependent youth; homeless youth; youth with lower socioeconomic status)

"Program effectiveness among high-school-aged youth (14-18 yrs)

"Program effectiveness in non-US settings

"Moderating effects of different program strategies, concepts, dosages, theoretical bases, formats, settings, facilitators, and contexts

"Moderating effects of implementation fidelity for existing programs

"Effects of strategies to improve the validity of self-report

Future research could consult the following methodological needs:

"Improved reporting of key methodological, clinical, and statistical information (e.g., method of randomization, allocation concealment, procedures for blinding data collectors, numbers of participants per trial arm at follow-up, attrition analyses, means and standard deviations for continuous outcomes). Use of the CONSORT statement (Moher 2001).

"Standardized behavioral outcome measures with consistent follow-up intervals, recall periods, and clinical meanings. Increased use of HIV, STI, and pregnancy as primary outcomes. Disaggregation of oral, anal, and vaginal sex acts. Use of medical evaluation for biological outcomes instead of self-report.

”More complete reporting of implementation data (i.e., program design, program delivery, participant uptake, and trial context) for intervention and control arms

“Analyses that account for dropouts (intention-to-treat) and unit of randomization

”Correction for multiple statistical tests

“Incorporation of cost-effectiveness and participant satisfaction data

”Provision of data sets for IPD (individual patient data) analyses

ACKNOWLEDGEMENTS

We thank Nandi Siegfried, Gail Kennedy, Tara Horvath, and George Rutherford for providing editorial guidance and ongoing support during the review process. We are grateful to Karishma Busgeeth for helping us develop our search strategy and for carrying out searches of CENTRAL, PubMed, AIDSLINE, and EMBASE in 2005. We thank the primary trialists who sent manuscripts where needed and responded to our queries about their primary trials. We also thank the many experts who responded to our queries regarding unpublished and ongoing evaluations, the trialists who helped us by sending manuscripts that were unpublished or impossible to obtain in the UK, and the statistical experts who provided us with guidance about how to deal with cluster-randomized trials. We also thank the Centre for Evidence-Based Intervention and the Department of Social Policy and Social Work at the University of Oxford.

REFERENCES

References to studies included in this review

Anderson 1999 *{published and unpublished data}*
Anderson NLR, Koniak-Griffin D, Keenan CK, Uman G, Duggal BR, Casey C. Evaluating the outcomes of parent-child family life education. *Sch Inq Nurs Pract* 1999;**13**(3): 211–234.

Blake 2001 *{published data only}*
Blake SM, Simkin L, Ledsky R, Perkins C, Calabrese JM. Effects of a parent-child communications intervention on young adolescents' risk for early onset of sexual intercourse. *Fam Plann Perspect* 2001;**33**(2):52–61.

Clark 2005 *{published data only}*
Clark LF, Miller KS, Nagy SS, et al.Adult identity mentoring: Reducing sexual risk for African-American seventh grade students. *J Adolesc Health* 2005;**37**(4): 337.e1–10.

Goldfarb 1999 *{published and unpublished data}*
Goldfarb E, Donnelly J, Duncan DF, Young M, Eadie C, Castiglia D. Evaluation of an abstinence-based curriculum for early adolescents: first year changes in sex attitudes, knowledge, and behavior. *North American Journal of Psychology* 1999;**1**(2):243–254.

Hernandez 1990 *{published data only}*
Hernandez JT, Smith FJ. Abstinence, protection, and decision-making: experimental trials on prototypic AIDS programs. *Health Educ Res* 1990;**5**(3):309–320.

Kirby 1997a *{published and unpublished data}*
Cagampang H, Barth RP, Korpi M, Kirby D. Education Now and Babies Later (ENABL): life history of a campaign to postpone sexual involvement. *Fam Plann Perspect* 1997; **29**:109–14.
Kirby D, Korpi M, Barth R, Cagampang H. *Evaluation of Education Now and Babies Later (ENABL): Final report*. Berkeley, CA: Family Welfare Research Group at the School of Social Welfare, University of California at Berkeley, and ETR Associates, 1995.
* Kirby D, Korpi M, Barth RP, Cagampang HH. The impact of the Postponing Sexual Involvement curriculum among youths in California. *Fam Plann Perspect* 1997;**29** (3):100–108.

Kirby 1997a peer-led *{published and unpublished data}*
Kirby D, Korpi M, Barth RP, Cagampang HH. The impact of the Postponing Sexual Involvement curriculum among youths in California. *Fam Plann Perspect* 1997;**29**(3): 100–108.

Kirby 1997a teachers *{published and unpublished data}*
* Kirby D, Korpi M, Barth RP, Cagampang HH. The impact of the Postponing Sexual Involvement curriculum among youths in California. *Fam Plann Perspect* 1997;**29** (3):100–108.

Kirby 1997b *{published and unpublished data}*
Cagampang H, Barth RP, Korpi M, Kirby D. Education Now and Babies Later (ENABL): life history of a campaign to postpone sexual involvement. *Fam Plann Perspect* 1997; **29**:109–14.
Kirby D, Korpi M, Barth R, Cagampang H. *Evaluation of Education Now and Babies Later (ENABL): Final report*. Berkeley, CA: Family Welfare Research Group at the School of Social Welfare, University of California at Berkeley, and ETR Associates, 1995.
* Kirby D, Korpi M, Barth RP, Cagampang HH. The impact of the Postponing Sexual Involvement curriculum among youths in California. *Fam Plann Perspect* 1997;**29** (3):100–108.

Kirby 1997c *{published and unpublished data}*
Cagampang H, Barth RP, Korpi M, Kirby D. Education Now and Babies Later (ENABL): life history of a campaign

to postpone sexual involvement. *Fam Plann Perspect* 1997; **29**:109–14.
Kirby D, Korpi M, Barth R, Cagampang H. *Evaluation of Education Now and Babies Later (ENABL): Final report*. Berkeley, CA: Family Welfare Research Group at the School of Social Welfare, University of California at Berkeley, and ETR Associates, 1995.
* Kirby D, Korpi M, Barth RP, Cagampang HH. The impact of the Postponing Sexual Involvement curriculum among youths in California. *Fam Plann Perspect* 1997;**29** (3):100–8.

Miller 1993 ***{published data only (unpublished sought but not used)}***
Miller BC, Norton MC, Jenson GO, Lee TR, Christopherson C, King PK. Impact evaluation of Facts and Feelings: a home-based video sex education curriculum. *Fam Relat* 1993;**42**(4):392–400.

Trenholm 2007a ***{unpublished data only}***
Devaney B, Johnson A, Maynard R, Trenholm C. *The evaluation of abstinence education programs funded under Title V, Section 510: Interim report*. Princeton, NJ: Mathematica Policy Research, Inc., 2002.
Maynard R, Trenholm C, Devaney B, Johnson A, Clark MA, Homrighausen J, Kalay E. *First year impacts of four Title V, Section 510 abstinence programs*. Princeton, NJ: Mathematica Policy Research, Inc., 2005.
* Trenholm C, Devaney B, Fortson K, Quay L, Wheeler J, Clark M. *Impacts of four Title V, Section 510 abstinence education programs: Final report*. Princeton, NJ: Mathematica Policy Research, Inc., 2007.

Trenholm 2007b ***{unpublished data only}***
Devaney B, Johnson A, Maynard R, Trenholm C. *The evaluation of abstinence education programs funded under Title V, Section 510: Interim report*. Princeton, NJ: Mathematica Policy Research, Inc., 2002.
Maynard R, Trenholm C, Devaney B, Johnson A, Clark MA, Homrighausen J, Kalay E. *First year impacts of four Title V, Section 510 abstinence programs*. Princeton, NJ: Mathematica Policy Research, Inc., 2005.
* Trenholm C, Devaney B, Fortson K, Quay L, Wheeler J, Clark M. *Impacts of four Title V, Section 510 abstinence education programs: Final report*. Princeton, NJ: Mathematica Policy Research, Inc., 2007.

Trenholm 2007c ***{unpublished data only}***
Devaney B, Johnson A, Maynard R, Trenholm C. *The evaluation of abstinence education programs funded under Title V, Section 510: Interim report*. Princeton, NJ: Mathematica Policy Research, Inc., 2002.
Maynard R, Trenholm C, Devaney B, Johnson A, Clark MA, Homrighausen J, Kalay E. *First year impacts of four Title V, Section 510 abstinence programs*. Princeton, NJ: Mathematica Policy Research, Inc., 2005.
* Trenholm C, Devaney B, Fortson K, Quay L, Wheeler J, Clark M. *Impacts of four Title V, Section 510 abstinence education programs: Final report*. Princeton, NJ: Mathematica Policy Research, Inc., 2007.

Trenholm 2007d ***{unpublished data only}***
Devaney B, Johnson A, Maynard R, Trenholm C. *The evaluation of abstinence education programs funded under Title V, Section 510: Interim report*. Princeton, NJ: Mathematica Policy Research, Inc., 2002.
Maynard R, Trenholm C, Devaney B, Johnson A, Clark MA, Homrighausen J, Kalay E. *First year impacts of four Title V, Section 510 abstinence programs*. Princeton, NJ: Mathematica Policy Research, Inc., 2005.
* Trenholm C, Devaney B, Fortson K, Quay L, Wheeler J, Clark M. *Impacts of four Title V, Section 510 abstinence education programs: Final report*. Princeton, NJ: Mathematica Policy Research, Inc., 2007.

References to studies excluded from this review

Abel 2007 ***{unpublished data only}***
Abel EM, Greco M. Assessing the effectiveness of an abstinence-based school/community intervention. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Barnett 2003 ***{published data only}***
Barnett JE, Hurst CS. Abstinence education for rural youth: an evaluation of the Life's Walk program. *J Sch Health* 2003;**73**(7):264–268.

Bearman 2001 ***{published data only}***
Bearman P, Bruckner H. Promising the future: virginity pledges and first intercourse. *American Journal of Sociology* 2001;**106**(4):859–912.

Bekenstein 2007 ***{unpublished data only}***
Bekenstein S, Garrity P, Posegate K, DuBois M, Goldberg B. Teens Taking Charge: A youth development approach to abstinence education. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Bersamin 2005 ***{published data only}***
Bersamin MM, Walker S, Waiters ED, Fisher DA, Grube JW. Promising to wait: virginity pledges and adolescent sexual behavior. *J Adolesc Health* 2005;**36**:428–436.

Borawski 2001 ***{published data only}***
* Borawski E, Lovegreen L, Demko C, Guwatudde D, Abbot K, Stewart S. *Evaluation of the teen pregnancy prevention programs funded through the Wellness Block Grant (1999-2000)*. Cleveland, OH: Center for Health Promotion and Research, Department of Epidemiology and Biostatistics, Case Western Reserve University, 2001.

Borawski 2005 ***{published data only}***
Borawski EA, Trapl ES, Lovegreen LD, Colabianchi N, Block T. Effectiveness of abstinence-only intervention in middle school teens. *Am J Health Behav* 2005;**29**(5): 423–34.

Bruckner 2005 ***{published data only}***
Bruckner H, Bearman P. After the promise: the STD consequences of adolescent virginity pledges. *J Adolesc Health* 2005;**36**:271–8.

Carter 2004 ***{unpublished data only}***
Carter T. Evaluation report for the Kansas Abstinence Education Program. Children and Families Section. Bureau

for Children, Youth and Families, Kansas Department of Health and Environment 2004.

Christopher 1990 *{published data only}*
Christopher FS, Roosa MW. An evaluation of an adolescent pregnancy prevention program: is "Just Say No" enough?. *Fam Relat* 1990;**39**:68–72.

Denny 1999 *{published data only}*
Denny G, Young M, Spear CE. An evaluation of the Sex Can Wait abstinence education curriculum series. *Am J Health Behav* 1999;**23**(2):134–143.

Denny 2002 *{published data only}*
Denny G, Young M. An evaluation of an abstinence-only sex education curriculum: an 18-month follow-up. *J Sch Health* 2006;**76**(8):414–22.
Denny G, Young M, Rausch S, Spear C. An evaluation of an abstinence education curriculum series: Sex Can Wait. *Am J Health Behav* 2002;**26**(5):366–377.

Denny 2003 *{published data only}*
Denny G, Young M. Evaluation of an abstinence education curriculum series: Sex Can Wait. Paper presented at the annual meeting of the American School Health Association. 2003.

DiFiore 2007 *{unpublished data only}*
DiFiore K, Ross S, Donnelly J, Mays V. Evaluation results from the Choice Game abstinence education program. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Doniger 2001 *{published data only}*
Doniger AS, Adams E, Utter CA, Riley JS. Impact evaluation of the "Not Me Not Now" abstinence-oriented, adolescent pregnancy prevention communications program, Monroe County, New York. *J Health Commun* 2001;**6**: 45–60.

Goodson 2004 *{published data only}*
Goodson P, Pruitt BE, Buhi E, Wilson KL, Rasberry CN, Gunnels E. Abstinence education evaluation phase 5 technical report. Department of Health & Kinesiology, Texas A&M University 2004.

Halpin 2007 *{unpublished data only}*
* Halpin G, Halpin G. Abstinence education: Program evaluation with youth survey. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.
Halpin G, Halpin G. Abstinence-only education: Validation and evaluation. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

Hubbard 1998 *{published data only}*
Hubbard BM, Giese ML, Rainey J. A replication study of reducing the risk, a theory-based sexuality curriculum for adolescents. *J Sch Health* 1998;**68**(6):243–247.

Janken 2005 *{unpublished data only}*
Janken JK, Minnich M, Sheppard BK, Canady M. Three-month evaluation results of TRAIL, an abstinence-based adolescent pregnancy prevention program. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

Jorgensen 1993 *{published data only}*
Jorgensen SR. Project Taking Charge: an evaluation of an adolescent pregnancy prevention program. *Fam Relat* 1991; **40**(4):373–380.
* Jorgenson SR, Potts V, Camp B. Project Taking Charge: six-month follow-up of a pregnancy prevention program for early adolescents. *Fam Relat* 1993;**42**(4):401–406.

Lage 2005 *{unpublished data only}*
Lage OG, Pratt IA, Claussen AH. The Multimedia Approach to Pregnancy Prevention (MAPP) Project for chronically ill youth. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

LeCroy 2003 *{published data only}*
LeCroy, Milligan Associates. Final report: Arizona abstinence only education program 1998-2003. Arizona Department of Health Services, Office of Women's and Children's Health 2003.

Lerner 2004 *{published data only}*
Lerner R. *Executive Summary: can abstinence work? An analysis of the Best Friends program*. Washington, DC: Best Friends Foundation, 2005.
Lerner R. Can abstinence work? An evaluation of the Best Friends program. *Adolescent and Family Health* 2004;**3**(4): 185–92.

Lieberman 2005 *{unpublished data only}*
Lieberman L, Berlin C, Skov N. The NiteStar StarLo Theater education approach to abstinence education for early adolescents in New York City. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

Lin 2007 *{unpublished data only}*
* Lin H-L, Vazin T. Changes in adolescents' behavioral intentions over time. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.
Lin H-L, Vazin T, Dimitrov DM. Effects of a program intervention on factors of teens' sexual activities. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Minnesota 2002 *{published data only}*
Professional Data Analysts Inc, Professional Evaluation Services. *Minnesota Education Now and Babies Later (MN ENABL): Evaluation Report 1998-2002*. St Paul, MN: Minnesota Department of Health, Division of Family Health, Maternal and Child Health Section, 2002.

Mitchell-DiCenso 97 *{published data only (unpublished sought but not used)}*
* Mitchell-DiCenso A, Thomas BH, Devlin MC, Goldsmith CH, Willan A, Singer J, Marks S, Watters D, Hewson S. Evaluation of an educational program to prevent adolescent pregnancy. *Health Educ Behav* 1997;**24**(3): 300–312.
Thomas BH, Mitchell A, Devlin MC, Goldsmith CH, Singer J, Watters D. Small group sex education at school: the McMaster Teen Program. In: Miller BC, Card JJ, Paikoff RL, Peterson JL editor(s). *Preventing Adolescent Pregnancy: Model Programs and Evaluations*. Newbury Park, CA: Sage Publications, 1992:28–52.

Nicholson 1992 *{published data only}*
Nicholson H, Postrado L. A comprehensive age-phased approach: Girls Incorporated. In: Miller BC, Card JJ, Paikoff RL, Peterson JL editor(s). *Preventing adolescent pregnancy: Model programs and evaluations*. Newbury Park, CA: Sage Publications, 1992:110–138.

O'Donnell 2005 *{published data only}*
O'Donnell L, Stueve A, Agronick G, Wilson SR, Duran R, Jeanbaptiste V. Saving sex for later: an evaluation of a parent education intervention. *Perspectives on Sexual and Reproductive Health* 2005;**37**(4):166–73.

Olsen 1991 *{published data only}*
Olsen JA, Weed SE. Effects of three abstinence sex education programs on student attitudes toward sexual activity. *Adolescence* 1991;**26**(103):631–41.

Olsen 1992 *{published data only}*
Olsen J, Weed S. Student evaluation of sex education programs advocating abstinence. *Adolescence* 1992;**27**(106): 369–80.

Pallone 2007 *{unpublished data only}*
Pallone K, Ferraro K. The PEERS Project's impact on abstinence commitments and rationale: An experimental design. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Pruitt 2005 *{published data only}*
Tanne JH. Abstinence-only programmes do not change sexual behavior, Texas study shows. *BMJ* 2005;**330**:326.

Resnick 1997 *{published data only}*
Resnick MD, Bearman PS, Blum RW, Bauman KE, Harris KM, Jones J, et al.Protecting adolescents from harm. Findings from the National Longitudinal Study on Adolescent Health. *JAMA* 1997;**278**(10):823–32.

Roberts-Gray 2007 *{unpublished data only}*
Roberts-Gray C. Evaluation of LifeWorks Adolescent Pregnancy Prevention Program. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.
* Roberts-Gray C, Pierce A, Lewis CM. Length of engagement time in abstinence education support groups and adolescent outcomes. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Roosa 1990 *{published data only}*
Roosa MW, Christopher FS. Evaluation of an abstinence-only adolescent pregnancy prevention program: A replication. *Fam Relat* 1990;**39**(4):363–7.

Rue 2005 *{unpublished data only}*
Rue L, Weed S. Primary prevention of adolescent risk-taking: A school-based model. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

Saunders 2005 *{published data only}*
Saunders EJ, Landsman MJ, Graf NM, Harms V. *Evaluation of abstinence only education in Iowa*. Iowa City, IA: Iowa Department of Public Health, 2005.

Smith 2003 *{published data only}*
Smith E, Dariotis J, Potter S. *Evaluation of the Pennsylvania Abstinence Education and Related Services initiative 1998-2002*. State College, PA: Prevention Research Center for the Promotion of Human Development, Pennsylvania State University, 2003.

Spear 1997 *{published data only}*
Spear C, Young M, Denny G. Field testing of an abstinence-based sexuality education program for upper elementary school students. *J Health Educ* 1997;**28**(6):335–340.

St Pierre 1995 *{published data only}*
St Pierre TL, Mark MM, Kaltreider DL, Aikin KJ. A 27-month evaluation of a sexual activity prevention program in Boys & Girls Clubs across the nation. *Fam Relat* 1995;**44** (1):69–77.

Sternas 2007 *{unpublished data only}*
Sternas K, Scharf MA, Peterkin R. Preliminary evaluation findings for the Newark, NJ Adolescent Family Life abstinence education program. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Sulak 2006 *{published and unpublished data}*
* Sulak P, Herbelin S, Kuehl A, Kuehl T, Fix DD. Analysis of knowledge and attitudes of young adolescents and adult groups particiating in an educational program regarding adolescent sex education with a focus on abstinence. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.
Sulak PJ, Herbelin SJ, Fix DDA, Kuehl TJ. Impact of an adolescent sex education program that was implemented by an academic medical center. *American Journal of Obstetrics and Gynecology* 2006;**195**(1):78–84.

Tanner 2007 *{unpublished data only}*
Tanner JF, Davis M. Examining the potential for decay: A 5-6 month post-test. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Tobin 2005 *{unpublished data only}*
Tobin T, Fuller R. Northwest Family Services' Youth Solutions: A comparison of brief and intensive abstinence education interventions. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

Usera 2007 *{unpublished data only}*
Usera J, Anagnopoulos C, Jenkins HE. The impact of a multifaceted pre- and early-adolescent female program on sexual abstinence and other risky behaviors. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.
* Usera JJ, Anagnopoulos C. The influence of moral competence on abstinence education for pre and early adolescent females. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Weed 1989 *{published data only}*
Weed SE, Olsen JA. *Evaluation Report of The Sex Respect Program: Results for the 1988-1989 School Year*. Salt Lake City, UT: Institute for Research and Evaluation, 1989.

Weed 1992 *{published data only}*

Olsen J, Weed S. Student evaluation of sex education programs advocating abstinence. *Adolescence* 1992;**27**(106): 369–80.

Olsen JA, Weed SE. Effects of three abstinence sex education programs on student attitudes toward sexual activity. *Adolescence* 1991;**26**(103):631–41.

* Weed SE, Olsen JA, DeGaston J, Prigmore J. *Predicting and Changing Teen Sexual Activity Rates: A Comparison of Three Title XX Programs*. Salt Lake City, UT: Institute for Research and Evaluation, 1992.

Weed SE, Prigmore J, Tanas R. *The Teen Aid Family Life Education Project: fifth year evaluation report*. Salt Lake City, UT: Institute for Research and Evaluation, 1992.

Weed 1994 *{published data only}*

Weed SE. *FACTS project: year end evaluation report 1993-1994*. Salt Lake City, UT: Institute for Research and Evaluation, 1994.

Weed 2001 *{published data only}*

Weed S. *Title V education programs: Phase I interim evaluation report to Arkansas Department of Health*. Salt Lake City, UT: Institute for research and evaluation, 2001.

Weed 2005a *{published data only}*

Weed SE, Ericksen IH, Birch PJ. *An evaluation of the Heritage Keepers abstinence education program*. Salt Lake City, UT: Institute for Research and Evaluation, 2005.

Weed 2005b *{published data only}*

Weed SE, Anderson N, Eriksen IH. *An evaluation of Choosing the Best*. Salt Lake City, UT: Institute for Research and Evaluation, 2005.

Welmaker 2005 *{unpublished data only}*

Welmaker R, Archie-Booker E, Langley M. Project Right CHOICES: A teenage pregnancy prevention program, three years later. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

Williams 2004 *{unpublished data only}*

Williams JD. Evaluating an abstinence-only program: is this an effective pregnancy and STD prevention method? . *Public Policy Working Paper Series, Public Policy Ph.D. Program*. Fayetteville, AR: University of Arkansas, 2004.

Young 1992 *{published data only}*

Young M. Abstinence-oriented sexuality education: initial field test results of the Living Smart curriculum. *Family Life Educator* 1992;**10**(4):4–8.

References to ongoing studies

Borawski 2007 *{unpublished data only}*

Borawski E, Landis C, Colabianchi N, Albert J. Research Projects: Healthy Teens Building Healthy Schools. Centre for Health Promotion Research, Case Western Reserve University 2007; Vol. http://www.case.edu/affil/healthpromotion/projects/healthy˙teens˙proj.html.

Jemmott 2006 *{unpublished data only}*

Jemmott JB, III, Jemmott LS, Fong GT. Efficacy of an abstinence-only intervention over 24 months: A randomized controlled trial with young adolescents. International AIDS Conference. Toronto, Canada, 2006.

Markham 2007 *{unpublished data only}*

Markham, C. SIP 4-04 Evaluation of Abstinence-Only and Abstinence-Plus Program to Prevent HIV, STD, and Pregnancy Among Middle School Students. *United States National Institute of Health* 2007;**http://clinicaltrials.gov/ct/show/NCT00167505?order=1**.

Miller-Heyl 2007 *{unpublished data only}*

MacPhee D, Podunovich R, Miller-Heyl J. Dare to Be You "Care to Wait" family-based intervention program. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

* Miller-Heyl J, Podunovich RA, MacPhee D. The Dare to be You "Care to Wait" Program: A multilevel, family-based, abstinence education research project. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

Miller-Heyl J, Washington D, Podunovich R. Adolescent Development: Under Pressure: Building Refusal Skills in Adolescents through Family-Based Intervention. Office of Adolescent Pregnancy Programs Prevention Grantee Conference. Pittsburgh, PA, 2006.

Nagel 2007 *{unpublished data only}*

Nagel R, Cameron C. Positive Choices: Pilot results and future plans. Abstinence Education Evaluation Conference. Baltimore, MD, 2005.

* Nagel R, Cameron C, Christopher C. Positive Choices abstinence education longitudinal program: Year 1 results. Abstinence Education Evaluation Conference. Baltimore, MD, 2007.

Additional references

Aarons 2000

Aarons S, Jenkins R, Raine T, et al.Postponing sexual intercourse among urban junior high school students: a randomized controlled evaluation. *J Adolesc Health* 2000; **27**:236–47.

Adams 2004

Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemi* 2004;**57**:785–94.

Alexander 1993

Alexander CS, Somerfield MR, Ensminger ME, Johnson KE, Kim YJ. Consistency of adolescents' self-report of sexual behavior in a longitudinal study. *Journal of Youth and Adolescence* 1993;**22**(5):455–71.

AVERT 2005

AVERT. UK HIV/AIDS FAQs. AVERT 2005; Vol. Retrieved 26 April 2005 from http://www.avert.org/aidsfaqs.htm.

Bailey 2002

Bailey W, Young M, Knickerbocker C, Doan T. A cautionary tale about conducting research on abstinence education: How do state abstinence coordinators define

"sexual activity"?. *American Journal of Health Education* 2002;**33**(5):290–296.

Bennett 2005

Bennet SE, Assefi NP. School-based teenage pregnancy prevention programs: a systematic review of randomized controlled trials. *J Adolesc Health* 2005;**36**:72–81.

Berne 1999

Berne L, Huberman B. *European approaches to adolescent sexual behavior and responsibility*. Washington DC: Advocates for Youth, 1999.

Berry 2005

Berry S. AIDS in America. AVERT 2005; Vol. Retrieved 26 April 2005 from http://www.avert.org/aids–america.htm.

Besharov 1997

Besharov DJ, Gardiner KN. Sex education and abstinence: Programs and evaluation. *Children and Youth Services Review* 1997;**19**(5/6):327–339.

Binson 1998

Binson D, Catania JA. Respondents' understanding of the words used in sexual behavior questions. *Public Opinion Quarterly* 1998;**62**:190–208.

Brener 2004

Brener N, Grunbaum J, Kann L, McManus T, Ross J. Assessing health risk behaviors among adolescents: the effect of question wording and appeals for honesty. *J Adolesc Health* 2004;**35**:91–100.

Brown 1997

Brown S. Sex education and abstinence programs: Why don't we know more?. *Children and Youth Services Review* 1997;**19**(5/6):327–339.

Cagampang 1997

Cagampang H, Barth RP, Korpi M, Kirby D. Education Now and Babies Later (ENABL): life history of a campaign to postpone sexual involvement. *Fam Plann Perspect* 1997; **29**:109–14.

Catania 1996

Catania JA, Binson D, Canchola J, Pollack LM, Hauck W. Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly* 1996;**60**(3):345–75.

Connolly 2004

Connolly C. Some abstinence programs mislead teens, report says. The Washington Post 2004, December 2; Vol. Retrieved 4 April 2005 from http://www.washingtonpost.com.

Crepaz 2002

Crepaz N, Marks G. Towards an understanding of sexual risk behavior in people living with HIV: A review of social, psychological, and medical findings. *AIDS* 2002;**16**: 135–149.

Devaney 2002

Devaney B, Johnson A, Maynard R, Trenholm C. *The evaluation of abstinence education programs funded under title V section 510: Interim report*. Princeton, NJ: Mathematica Policy Research, Inc., 2002.

DiCenso 2002

DiCenso A, Guyatt G, Willan A, Griffith L. Interventions to reduce unintended pregnancies among adolescents: Systematic review of randomised controlled trials. *British Medical Journal* 2002;**324**:1426–1435.

Donner 2002

Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Statistics in Medicine* 2002;**21**:2971–80.

Egger 1997

Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997;**315**(7109):629–34.

Farmer 2001

Farmer P. *Infections and Inequalities: The Modern Plagues*. Berkeley, CA: University of California Press, 2001.

Franklin 1997

Franklin C, Grant D, Corcoran J, Miller PO, Bultman L. Effectiveness of prevention programs for adolescent pregnancy: a meta-analysis. *Journal of Marriage and the Family* 1997;**59**(3):551–67.

Grimwade 2003

Grimwade K, Swingler G. Cotrimoxazole prophylaxis for opportunistic infections in adults with HIV. *Cochrane Database of Systematic Reviews 2003* 2003, Issue 3, Art. No.: CD003108. DOI: 10.1002/14651858.CD003108.[Art. No.: CD003108. DOI: 10.1002/14651858.CD003108]

Grunseit 1997

Grunseit A, Kippax S, Aggleton P, Baldo M, Slutkin G. Sexuality education and young people's sexual behavior: a review of studies. *J Adolesc Res* 1997;**12**(4):421–53.

Haglund 2002

Haglund K. Sexually abstinent African American adolescent females' descriptions of abstinence. *Journal of Nursing Scholarship* 2002;**35**(3):231–236.

Haskins 1997

Haskins R, Bevan CS. Abstinence education under welfare reform. *Children and Youth Services Review* 1997;**19**(5/6): 465–484.

Higgins 2002

Higgins J, Thompson S. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* 2002;**21**(11):1538–58.

Higgins 2005

Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5. *The Cochrane Library, Issue 3*. John Wiley & Sons, Ltd, 2005.

Hollis 1999

Hollis S, Campbell F. What is meant by intention-to-treat analysis?. *BMJ* 1999;**319**:670–4.

Horan 1998

Horan P, Phillips J, Hagen N. The meaning of abstinence for college students. *Journal of HIV/AIDS Prevention and Education for Adolescents and Children* 1998;**2**:51–66.

Howard 1990

Howard M, McCabe J. Helping teenagers postpone sexual involvement. *Fam Plann Perspect* 1990;**22**:21–6.

Howard 1992
Howard M, McCabe JA. An information and skills approach for younger teens: Postponing Sexual Involvement program. In: Miller BC, Card JJ, Paikoff RL, Peterson JL editor(s). *Preventing Adolescent Pregnancy: Model Programs and Evaluations*. Sage Publications, 1992:83–109.

HRW 2002
Human Rights Watch. *Ignorance only: HIV/AIDS, human rights, and abstinence-only programs in the United States*. New York, NY: Human Rights Watch, 2002.

Jemmott 1992
Jemmott JB, III, Jemmott LS, Fong GT. Reductions in HIV risk-associated sexual behaviors among Black male adolescents: Effects of an AIDS prevention intervention. *Am J Public Health* 1992;**82**(3):372–7.

Jemmott 1998
Jemmott JB III, Jemmott LS, Fong GT. Abstinence and safer sex HIV risk-reduction interventions for African American adolescents: a randomized controlled trial. *JAMA* 1998;**279**(19):1529–36.

Jemmott 1999
Jemmott JB, III, 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**(2): 161–87.

Jemmott 2000
Jemmott JB III, Jemmott LS. HIV risk reduction behavioral interventions with heterosexual adolescents. *AIDS* 2000;**14** (Supplement 2):S40–S52.

Jemmott 2004
Jemmott JB, III, Jemmott LS, Fong GT, Hines PM. Evaluation of an HIV/STD risk reduction intervention implemented by Non-Governmental Organizations (NGOs): A randomized controlled cluster trial. International AIDS Conference. Bangkok, Thailand, 2004.

Johnson 2002
Johnson W, Semaan S, Hedges L, Ramirez G, Mullen P, Sogolow E. A protocol for the analytical aspects of a systematic review of HIV prevention research. *J Acquir Immune Defic Syndr* 2002;**30**:S62–72.

Jones 1985
Jones E, Forrest J, Goldman N, Henshaw S, Lincoln R, Rosoff J, et al.Teenage pregnancy in developed countries: Determinants and policy implications. *Fam Plann Perspect* 1985;**17**(2):53–63.

Kim 1997
Kim N, Stanton B, Li X, Dickersin K, Galbraith J. Effectiveness of the 40 adolescent AIDS-risk reduction interventions: a quantitative review. *J Adolesc Health* 1997; **20**:204–215.

Kirby 1995
Kirby D, Korpi M, Barth R, Cagampang H. *Evaluation of Education Now and Babies Later (ENABL): Final report*. Berkeley, CA: Family Welfare Research Group at the School of Social Welfare, University of California at Berkeley, and ETR Associates, 1995.

Kirby 1997R
Kirby D, Coyle K. School-based programs to reduce sexual risk-taking behavior. *Children and Youth Services Review* 1997;**19**(5/6):415–36.

Kirby 2001
Kirby D. *Emerging answers: Research findings on programs to reduce teen pregnancy*. Washington DC: National Campaign to Prevent Teen Pregnancy, 2001.

Kirby 2002
Kirby D. *Do abstinence-only programs delay the initiation of sex among young people and reduce teen pregnancy?*. Washington, DC: National Campaign to Prevent Teen Pregnancy, 2002.

Kirby 2006
Kirby D, Laris BA, Rolleri L. *The impact of sex and HIV education programs in schools and communities on sexual behaviors among young adults*. Research Triangle Park, NC: Family Health International, YouthNet Program, 2006.

Lancet 2004
Editorial. Is it churlish to criticize Bush over his spending on AIDS?. *Lancet* 2004;**364**(9431):303–4.

Lauritsen 1997
Lauritsen JL, Swicegood CG. The consistency of self-reported initiation of sexual activity. *Fam Plann Perspect* 1997;**29**(5):215–21.

Manlove 2004
Manlove J, Papillo AR, Ikramullah E. *Not yet: programs to delay first sex among teens*. Washington, DC: National Campaign to Prevent Teen Pregnancy, 2004.

Maynard 2005
Maynard R, Trenholm C, Devaney B, Johnson A, Clark MA, Homrighausen J, Kalay E. *First year impacts of four Title V, Section 510 abstinence programs*. Princeton, NJ: Mathematica Policy Research, Inc., 2005.

Meston 1998
Meston CM, Heiman JR, Trapnell PD, Paulhus DL. Socially desirable responding and sexuality self-reports. *Journal of Sex Research* 1998;**35**(2):147–57.

Moher 2001
Moher D, Schulz KF, Altman D for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;**357**(9263):1191–4.

Montgomery 2005
Montgomery P, Gardner F, Operario D, Mayo-Wilson E, Tamayo S, Underhill K. The Oxford implementation Reporting Index: the development of an indicator of treatment fidelity in systematic reviews of psycho-social interventions (Poster). Cochrane Collaboration Colloquium. Melbourne, Australia, 2005.

Moore 1997
Moore K, Sugland B. Using behavioral theories to design abstinence programs. *Children and Youth Services Review* 1997;**19**(5/6):485–500.

Morris 2006
Morris M, Handcock M, Miller W, et al.Prevalence of HIV infection among young adults in the United States: results from the Add Health study. *Am J Public Health* 2006;**96**(6): 1091–7.

Napier 1997
Napier K. Chastity programs shatter sex-ed myths. *Policy Review* 1997;**May/June**:12–23.

Newcomer 1988
Newcomer S, Udry JR. Adolescents' honesty in a survey of sexual behavior. *J Adolesc Res* 1988;**3**(3/4):419–23.

NHS 1997
NHS Centre for Reviews and Dissemination. Preventing and reducing the adverse effects of unintended teenage pregnancies. *Effective Health Care* 1997;**3**(1):1–12.

O'Leary 1997
O'Leary A, DiClemente RJ, Aral SO. Reflections on the design and reporting of STD/HIV behavioral intervention research. *AIDS Educ Prev* 1997;**9**(Supplement 1):1–14.

O'Reilly 2004
O'Reilly K, Medley A, Dennison J, Schmid GP, Sweat MD. Systematic review of the impact of abstinence-only programmes on risk behavior in developing countries. eJIAS. 2004; Vol. 1(1):TuPeC4899.

O'Reilly 2006
O'Reilly KR, Medley A, Dennison J, Sweat MD. Systematic review of the impact of abstinence-only programmes on risk behavior in developing countries (1990-2005). International AIDS Conference. Toronto, Canada, 2006.

OMB 2006
Office of Management and Budget. *Budget of the United States Government, FY 2007: Health and Human Services*. Washington, DC: Executive Office of the President of the United States, http://www.whitehouse.gov/omb/budget/fy2007/hhs.html, 2006.

Pedlow 2003
Pedlow C, Carey M. HIV sexual risk-reduction interventions for youth: A review and methodological critique of randomized controlled trials. *Behavior Modification* 2003; **27**:135–190.

Peterman 2000
Peterman TA, Lin LS, Newman DR, et al.Does measured behavior reflect STD risk? An analysis of data from a randomized controlled behavioral intervention study. *Sex Transm Dis* 2000;**27**(8):446–51.

Phan 2004
Phan K. Conservatives: Waxman report is faulty, not abstinence-only education. The Christian Post 2004, December 7; Vol. Retrieved 4 April 2005 from http://www.christianpost.com.

Pitts 2001
Pitts M, Rahman Q. Which behaviors constitute "having sex" among university students in the UK?. *Archives of Sexual Behavior* 2001;**30**(2):169–76.

Rector 2002
Rector R. *The effectiveness of abstinence education programs in reducing sexual activity among youth*. Washington DC: Heritage Foundation, 2002.

Remez 2000
Remez L. Oral sex among adolescents: Is it sex or is it abstinence?. *Family Planning Perspectives* 2000;**32**(6): 298–304.

Rivers 2000
Rivers K, Aggleton P. HIV prevention in industrialized countries. In: Peterson J, DiClemente R editor(s). *HIV Prevention Handbook*. London: Kluwer Academic/Plenum Publishers, 2000.

Robin 2004
Robin L, Dittus P, Whitaker D, et al.Behavioral interventions to reduce incidence of HIV, STD, and pregnancy among adolescents: A decade in review. *Journal of Adolescent Health* 2004;**34**:3–26.

Rosenbaum 2006
Rosenbaum J. Reborn a virgin: adolescents' retracting of virginity pledges and sexual histories. *Am J Public Health* 2006;**96**(6):1098–103.

Rotheram-Borus 2000
Rotheram-Borus MJ, O'Keefe Z, Kracker R, Foo HH. Prevention of HIV among adolescents. *Prevention Science* 2000;**1**(1):15–30.

Sanders 1999
Sanders S, Reinisch J. Would you say you "had sex" if...?. *Journal of the American Medical Association* 1999;**281**(3): 275–277.

Saunders 1996
Saunders DJ. Wilson abstains from costly ENABLERS. San Francisco Chronicle 17 December 1996.

Siegfried 2003
Siegfried N, Muller M, Volmink J, Deeks J, Egger M, Low N, et al.Male circumcision for prevention of heterosexual acquisition of HIV in men. *Cochrane Database of Systematic Reviews 2003* 2003, Issue 3, Art. No.: CD003362. DOI: 10.1002/14651858.CD003362.[Art. No.: CD003362. DOI: 10.1002/14651858.CD003362.pub2]

Silva 2002
Silva M. The effectiveness of school-based sex education programs in the promotion of abstinent behavior: a meta-analysis. *Health Educ Res* 2002;**17**(4):471–81.

Smoak 2006
Smoak ND, Scott-Sheldon LAJ, Johnson B, Carey MP, and the SHARP Research Team. Sexual risk reduction interventions do not inadvertently increase the overall frequency of sexual behavior: a meta-analysis of 174 studies with 116,735 participants. *J Acquir Immune Defic Syndr* 2006;**41**(3):374–84.

Sonfield 2001
Sonfield A, Gold RB. States' implementation of the Section 510 abstinence education program, FY 1999. *Family Planning Perspectives* 2001;**33**(4):166–171.

Thomas 2000
Thomas M. Abstinence-Based Programs for Prevention of Adolescent Pregnancies. *Journal of Adolescent Health* 2000; **26**:5–17.

UNAIDS 2004
UNAIDS/WHO. *2004 Report on the Global AIDS Epidemic: 4th Global Report*. Geneva: UNAIDS, 2004.

UNAIDS 2006
UNAIDS. *Report on the global AIDS epidemic*. Geneva, Switzerland: UNAIDS/WHO, 2006.

US 2003
United States Congress. United States Leadership Against HIV/AIDS, Tuberculosis, and Malaria. 22 U.S.C. 7672 2003.

Waxman 2004
Waxman HA. *The content of federally funded abstinence-only education programs*. Washington DC: United States House of Representatives, Committee on Government Reform - Minority Staff, 2004.

Wilson 2005
Wilson KL, Goodson P, Pruitt BE, Buhi E, Davis-Gunnels E. A review of 21 curricula for abstinence-only-until-marriage programs. *J Sch Health* 2005;**75**(3):90–8.

World Bank 2007
World Bank. High Income Economies. World Bank Vol. Retrieved 9 April 2007 from http://www.worldbank.org.

Young 1998
Young M, Bailey W. The politics of abstinence education: From Reagan's Title XX to welfare reform's Title V. *The International Electronic Journal of Health Education* 1998;**1**: 207–212.

Zabin 1984
Zabin LS, Hirsch MB, Smith EA, Hardy JB. Adolescent sexual attitudes and behavior: are they consistent?. *Fam Plann Perspect* 1984;**16**(4):181–5.

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Anderson 1999

Methods	RCT, Abstinence-only vs. Nonintervention (delayed intervention group) Unit of randomization: "Naturally occurring group" (e.g., after-school programs were randomized by school) Method of randomization: Not reported. Distribution to treatment arms was unequal: 295 students were in the intervention group, and 110 were in the control group. Allocation concealment: B (Unclear). Data collection: Anonymous English or Spanish self-report survey completed in program setting Assessment times: Baseline, immediate post-intervention (unclear duration from baseline), and 12m Complete case analysis at 12m follow-up; intention-to-treat analysis for immediate post-intervention follow-up due to 0% attrition Attrition: 0% at immediate posttest, 38.0% at 12m. Dropouts were older by one year, more likely to be male, and less likely to report skipping school than those retained at 12m follow-up. Group equivalence: No baseline differences in demographics, "intervening, variables" or risk behavior variables
Participants	Los Angeles County, CA (USA) N = 405 participants at baseline. Unclear number of clusters. Mean age: 10.6 yrs, range 9-14 yrs. Primarily 5th and 6th grade students Ethnicity: of the 251 students providing data at 12m follow-up, 2.4% Native American, 5.6% Asian American, 20.7% African American, 45.8% Hispanic, 12.7% Other White, 5.2% Mixed/Other, 7.6% No Response SES: Unclear. of the 251 students providing data at 12m follow-up, over 60% lived with both parents, and over 80% reported being college bound. Mothers' education averaged between high school graduate and some college coursework 40.2% Male, of the 251 participants providing data at 12m follow-up Notes on recruitment: Families were recruited during community meetings announced by schools and community agencies
Interventions	1: "Reaching Adolescents and Parents," developed by the American Red Cross. an 8-session curriculum included 6 sessions for adolescents, 1 session for adolescents and parents, and 1 session for parents. Unclear how many weeks elapsed during the program. Goals were to increase student knowledge about puberty and reproduction, improve communication and decision-making skills, facilitate family communication, and prevent STIs and pregnancy by delaying the onset of sexual activity. Activities included small group discussions, video clips, question-and-answer sessions, discussion of media and peer pressure, confidence-building skill activities, and homework assignments. Theoretical basis: Cognitive-behavioral therapy, social learning theory. Setting: Community centers and schools. Exposure: 8 sessions, duration of sessions unclear. Staff: Trained teachers. 2: Delayed intervention: RAP was delivered after the evaluation was completed. No intervention or activities were provided during the course of the evaluation
Outcomes	Ever been pregnant or ever caused a pregnancy among all participants, excluding pregnancies reported at baseline (4 in intervention group, 1 in control group)

Anderson 1999 (*Continued*)

Notes	Funding: US Department of Health and Human Services, Office of Adolescent Pregnancy Programs The study originally collected data on the number of times participants had intercourse over the past 3m, but this was removed from surveys due to parental objections Disparity between group sizes. Study was not designed to track individual participants. The abstract describes this as a "quasi-experimental evaluation," but participant groups were randomly assigned to treatment conditions

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Blake 2001

Methods	RCT, Abstinence-only enhanced (with parent-child homework activities) vs. Abstinence-only non-enhanced (without parent-child activities) Unit of randomization: Classroom (stratified by quarter marking period between schools) Method of randomization: Not reported. Allocation concealment: B (Unclear). Assessment by confidential self-report survey completed in class Assessment times: Baseline and immediately (< 1 week) post-intervention. 7 weeks elapsed between baseline and follow-up assessment Complete case analysis. Attrition: 9.8% at 7 wks. Attrition analyses not provided. Baseline differences in belief that substance use increases risk, belief that media influences sexual behavior, % going "further sexually than wanted" in previous 3m, likelihood of being in high-risk sexual situations in past 3m, % reporting lifetime alcohol use, % reporting recent alcohol use. Controlled for baseline values in analyses
Participants	Rochester, NY (USA) N = 389 participants at baseline. Clusters were 19 classrooms in 3 schools Mean age: Specific age not reported. Participants were in 8th grade (approx 13.5 yrs) Ethnicity: 85% Caucasian (non-Hispanic). Others not specified SES: Predominantly middle-class suburban communities. 52% Male, of the 351 participants with complete case data. Notes on recruitment: Entire schools were recruited, and included classrooms were drawn from 8th grade health or consumer science classes
Interventions	1: "Managing the Pressures Before Marriage (MPM) - Enhanced," based on the Postponing Sexual Involvement curriculum. 5 class sessions over 5 weeks were taught by slightly older peer leaders recruited from local high schools. Sessions focused on understanding the risks of early sex, limit-setting, and assertiveness training. the program stressed abstinence specifically until marriage. Classroom activities included brainstorming, roleplays, skills training and rehearsal. Participants were assigned an additional 5 homework assignments (each with 3-5 activities) to complete with their parents. These assignments focused on understanding the pressures for early sex and improving parent-child communication. Homework completion was monitored.

Blake 2001 (*Continued*)

	Theoretical basis: Social learning theory, social cognitive theory. Setting: Curriculum took place in middle school classrooms, and homework was completed outside of school. Exposure: 5 class periods over 5 weeks, plus 5 homework assignments. Staff: Pairs of youth leaders from local high schools. 2: MPM- non-enhanced. This intervention was the same as in condition 1, but participants were not assigned the 5 additional parent-child homework assignments. All other characteristics the same as 1. the same youth leaders facilitated both sets of programs, but they were not told why control group classes did not receive homework assignments
Outcomes	Sexual intercourse ever among all participants Sexual intercourse in past 3m among all participants Lifetime number of partners - not reported due to the small number of students reporting sexual intercourse Regularity of condom use - not reported due to the small number of students reporting sexual intercourse
Notes	Funding: US Department of Health and Human Services, Office of Adolescent Pregnancy Programs Random assignment resulted in unequal distribution: 11 intervention classrooms, 8 control classrooms Follow-up tests were administered at 7 weeks after baseline, but the recall period for "recent sexual intercourse" was 3 months. These data may therefore span the time period before the intervention was administered Very short follow-up. Self-selection evident in the percentage of condition 1 students choosing to complete the homework: Non-completion of homework assignments in the MPM-enhanced group was significantly higher among males, students reporting recent sexual intercourse, students who did not receive "mostly As in school," and black or Hispanic students (as opposed to non-Hispanic whites) Fidelity monitoring: a program staff member attended each intervention session, and the trial report states that there was no evidence that the youth leaders' program implementation differed systematically between the intervention and control groups

Risk of bias

Item	**Authors' judgement**	**Description**
Allocation concealment?	Unclear	B - Unclear

Clark 2005

Methods	RCT, Abstinence-only vs. Usual Care. Unit of randomization: Classroom. Method of randomization: Not reported. Allocation concealment: B (Unclear). Assessment by written self-report survey completed in health education classroom at baseline, and in various locations (e.g., library, auditorium, classrooms) Assessment times: baseline, 4.4 months, 12 months. Complete case analysis. Attrition: 15% at 4.4 months, 37% at 12 months. No differences in attrition by group or any other characteristic Baseline differences in sexual activity favoring the control group. These were controlled in analysis
Participants	"Southeastern metropolitan area" (USA) N = 248 youth randomized. Clusters were 20 classrooms in 1 school Mean age: 12.6. Ethnicity: 98% African-American. SES low. 55% male. Notes on recruitment: Eligible participants were African-American 7th-grade students who were enrolled in a required 9-week health education class
Interventions	1: Adult Identity Mentoring Project. 10 classes over 6 weeks were delivered in 7th grade classrooms. the program encouraged participants to envision their future selves, and to think about how their choices (risk and protective behaviors) would affect future goals. Concepts included legacies, occupations, marketplace skills, creating goals, and motivation to avoid risky behaviors. Activities included interacting with visiting role models, creating business cards, writing business letters, creating resumes, learning relationship skills, establishing a goal timeline, and small-group activities. Theoretical basis: Theory of possible selves. Setting: 7th-grade school health education classrooms. Exposure: 10 sessions over 6 weeks. Staff: A male-female pair of African-American college graduate students 2: Usual care health education, as determined by the school and taught by regular classroom health educators
Outcomes	Sex since baseline among all participants. Sex since baseline among participants who reported never having had sex at baseline
Notes	Funding: Centers for Disease Control and Prevention, National Institute of Health Participants received coupons for $5 for each completed assessment No information about the monitoring of program implementation 11 classrooms were assigned to the AIM program, 9 classrooms were assigned to usual care

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Goldfarb 1999

Methods	RCT, Abstinence-only vs. Usual Care. Unit of randomization: School, stratified by district. Randomization was conducted according to matched-pair design based on school demographics Method of randomization: Not reported. the trial report does not state that randomization was used; this was discovered via personal communication with the trialist Allocation concealment: B (Unclear). Assessment by anonymous self-report surveys completed in school. Surveys were administered by project staff at baseline, and were administered by the students' teachers at posttest Assessment times: Baseline, immediately after the intervention (8-12 weeks after baseline) and at the following school year. Analyses for the final follow-up were unavailable, and it was unclear how many months elapsed before this last assessment took place Complete case analysis. Attrition unclear, because the report does not state the number of youth present at baseline. Analyses on 839 students providing data at post-test Groups were "comparable" at baseline, although specific data were not provided. At posttest, significant differences were observed between groups in age distribution (fewer control subjects in the oldest age group and highest grade) and housing (majority of intervention students lived in house while majority of controls lived in apartment). No differences by other demographics or religiosity
Participants	Northern New Jersey (USA) N = Baseline enrollment not reported. Analyses represented 839 students, 6 schools, 3 districts Age at baseline: Specific age not reported. Participants were in 6th through 8th grade (approximately 12. 5 yrs) Ethnicity: Communities described as "high minority." SES: Communities described as "lower socioeconomic status." 47.8% Male, of the 839 participants with complete case data. Notes on recruitment: Recruitment by school, 3 urban districts. Intervention delivered to entire school, with different programs for 6th, 7th, and 8th grades, although all three grades were grouped for analysis
Interventions	1: "Project C.A.R.E. (Creating Awareness through Relationship Education)." Separate curricula were delivered to 6th, 7th, and 8th grade students (although the three groups were analyzed together). 6th grade students received the 23 upper elementary lessons from the "Sex Can Wait" curriculum. 7th grade students received the 23 middle school lessons from the "Sex Can Wait" curriculum. 8th grade students received 23 of 45 stand-alone program activities from the "Abstinence Pick and Choose Activities" curriculum. All three sequences focused on (a) reproductive anatomy, physiology, and puberty; (b) communication skills; and (c) goal setting and life planning. All curricula targeted susceptibility, self-efficacy, social support, peer norms, factual information, and skill building. Parent-child homework assignments were included. Theoretical basis: Social learning theory. Setting: Public junior high school classrooms. Exposure: 8-12 weeks, involving 23 lessons for 6th graders, 24 lessons for 7th graders, and 23 lessons for 8th graders. Staff: Middle school teachers. 2: Usual Care - health curricula already in place, as dictated by the school
Outcomes	Ever had sex among all participants Sex in the last month among all participants Sex without a condom in the last month among all participants "Used a condom." It was unclear whether this referred to condom use in the last month among participants reporting sex in the last month, or condom use ever among participants reporting sex ever

Goldfarb 1999 (*Continued*)

Notes	Funding: US Department of Health and Human Services, Office of Adolescent Pregnancy Programs Control group and intervention group may have had the follow-up assessment at incomparable intervals; the intervention group received a post-test assessment "approximately 8-12 weeks" after baseline, but the intervention group received this assessment "upon completion" of the usual-care curriculum administered to them Teachers administered surveys to their own classes, which may have affected self-reported data Assessment at an additional later follow-up ("a post-post test survey") was conducted, but data were unavailable Study was not designed to track individual participants. Baseline enrollment and attrition rates not established. Exact time between baseline and final assessment unclear. Some student self-reports were inconsistent, e.g., a higher percentage of controls reported "recent sex" than reported "sex ever."

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Hernandez 1990

Methods	RCT, Abstinence-only vs. Abstinence-plus vs. Safer-sex vs. Nonintervention Unit of randomization: Individual. Method of randomization: Not reported. Allocation concealment: B (Unclear). Assessment via confidential written questionnaire in a group setting Assessment times: Baseline and 6 week follow-up. Complete case analysis. Attrition 5% for entire 4-arm study at 6 weeks. Attrition analyses not provided No statistically significant baseline differences.
Participants	North Carolina (USA) N = 410 at baseline. Mean age: 19.25 yrs. Ethnicity: 85% white. SES: Not reported. 55% Male. Notes on recruitment: 410 unmarried students aged 18-21 from state universities were recruited from an introductory psychology course. Participation in the study was an option to fulfill research requirements in the course
Interventions	1: Abstinence-only intervention ("Abstinence"). One session. Concepts included abstinence, the biology and symptoms of AIDS, transmission, and prevention. Participants received a 7-page written health education program (adapted from the North Carolina Board of Education senior high school AIDS curriculum); they then watched the video "AIDS: What everyone should know" (from Aims Media in Los Angeles). They also received a brochure about AIDS transmission and prevention (from the Fort Bragg Dependents' School System).

	Theoretical basis: Not reported. Setting: College classroom. Exposure: 45 minutes. Staff: Media-based, unclear what staff facilitated the program 2: Abstinence-plus intervention ("Protection"). Format, staff, exposure, theory, and setting as above, but the intervention "presented abstinence as the safest alternative and safe sex as another alternative." Concepts also included explicit instructions on condom use and discussing safe sex with partners. the written education program amalgamated brochures written by the Duke University Student Health Service and the American College Health Association ("Making Sex Safer"). the video was "AIDS: You're Not Immune" from the Alpha Theta fraternity. the brochure was from HERO, entitled "Safer Sex for Men and Women Concerned About AIDS." 3: Safe-sex intervention ("Decision"). Format, staff, exposure, theory, and setting as above, but the intervention "presented safe sex and described it as an alternative to abstinence." the written curriculum was adapted from the American Red Cross AIDS program for senior high school students. the video was "Letter from Brian," from the American Red Cross. the brochure "AIDS: the Facts" was also from the American Red Cross 4: Nonintervention control.
Outcomes	Number of times used condoms in the past 6 weeks, among all participants Number of times had sex in the past 6 weeks, among all participants Number of sex partners in the past 6 weeks, among all participants
Notes	Funding: not reported. This was the only study to enroll participants older than middle-school-age Condom use outcome measures the number of times participants used condoms, not the proportion of sexual acts that were protected Very short-term follow-up. Analysis was by 4x2x2x2 factorial design (program x gender x baseline sexual activity x time), analyses presented as "rank ordered interaction means." A test statistic was only reported for one outcome, and no post-hoc tests were provided

Risk of bias

Item	**Authors' judgement**	**Description**
Allocation concealment?	Unclear	B - Unclear

Kirby 1997a

Methods	RCT, Abstinence-only vs. Usual Care. Unit of randomization: Classroom, stratified by school. In 18 schools, participants were randomized to peer-led intervention, adult-led intervention, or usual care control. In 7 schools, participants were randomized to adult-led intervention or usual care control. Outcomes for the peer-led and adult-led groups were analyzed separately, but could not be separated into 2 separate designs for this review due to the overlap in control groups Method of randomization: Not reported. Allocation concealment: B (Unclear). Assessment by self-report surveys in English and Spanish, completed in school. Assessment times: Baseline, 3m, and 17m. Complete case analysis. Attrition: 17.6% at 3m, 33.7% at 17m. No significant differences among dropouts from intervention and control groups in all 3 designs; approx 1% more were lost from the intervention than from the control group in all 3 designs No significant baseline differences. Controlled for baseline values in analyses
Participants	California (USA) N = 4652 participants at baseline. At 3-month follow-up, clusters were 285 classrooms (of which 230 had been randomized between peer-led, adult-led, and usual care; 55 had been randomized between adult-led and usual care). At 17-month follow-up, clusters were 276 classrooms (223 had been randomized among all 3 conditions, and 53 had been randomized between adult-led and usual care) Mean age: 12.9 for peer-led comparisons, 12.9 for adult-led comparisons Ethnicity: for peer-led comparisons, 3.7% American Indian, 7.9% Asian/Pacific Islander, 10.0% African-American, 47.8% Hispanic, 20.9% White, 7.4% Other (percentages do not add to 100%). for adult-led comparisons, respective percentages were 4.6%, 8.7%, 7.5%, 44.5%, 27.9%, 8.0% (percentages do not add to 100%) SES: Not reported. % Male: for peer-led comparisons, 43.5% male. for adult-led comparisons, 41.5% male Notes on recruitment: Contractors, including schools and community-based organizations, applied to the California Office of Family Planning for funding to implement PSI and ENABL. They were selected because they served communities with high teen birthrates and diversity, and they demonstrated the ability to deliver PSI
Interventions	1: "Postponing Sexual Involvement (PSI)" delivered as part of the "Education Now and Babies Later (ENABL)" initiative. PSI was implemented in addition to whatever standard sexuality curriculum individual schools offered. Students received instruction in human sexuality, and then received a 5-session curriculum in class delivered by either adult facilitators or peer leaders. Concepts included risks of early sex, reasons to have sex or wait, resisting social pressures, limit-setting, assertive responses to resist sexual pressures. Activities included class discussions, group activities, videos, slides, and roleplays. Theoretical basis: Not reported, but original program was designed using social inoculation theory. Setting: Middle school classes (courses included science, health, physical education, and others). Exposure: 5 classes, 45-60 mins long. 4 sessions were designed for delivery over 4 consecutive weeks, with the 5th session delivered between one week and one month later. Peer staff: Pairs of peer leaders (peer leaders were in high school, slightly older than the study participants) . Peer leaders were always accompanied by adult leaders. Adult staff: Professional educators or college interns. Approximately 40% of adult leaders were health educators, 23% were classroom teachers, 23% were student interns, and the others were nurses, youth workers, or volunteers 2: Usual care - standard sexuality curriculum offered by school (varied). Generally was not comparable to PSI. Instead of PSI, youth typically received instruction in another topic area

Kirby 1997a (*Continued*)

Outcomes	STD diagnosis among all participants Pregnancy incidence among all participants Mean change in number of acts of intercourse in last 3m among participants reporting sexual experience at baseline or follow-up Mean change in number of acts of intercourse in last 12m among participants reporting sexual experience at baseline or follow-up Mean change in number of partners ever among participants reporting sexual experience at baseline or follow-up Condom use at last act of sexual intercourse among participants reporting sexual experience at baseline or follow-up Ever had sex among participants who were sexually inexperienced at baseline
Notes	Funding: Main funding from the California Office of Family Planning. Additional support from the Cowell Foundation, the Packard Foundation, and the Stuart Foundations Part of 3-design study. In other designs, youth were randomized by school and on an individual basis to either the adult-led PSI program or controls Extensive reporting on implementation challenges and participant satisfaction. PSI was delivered in the context of ENABL, a large statewide media and school-based campaign intended to raise awareness of teen pregnancy, encourage involvement in ENABL initiatives, and encourage a more supportive environment for youth sexual abstinence. Program contractors often established "community coalitions" and conducted community awareness activities

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Kirby 1997a peer-led

Methods	For purposes of analyses only
Participants	
Interventions	
Outcomes	
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Kirby 1997a teachers

Methods	For purposes of analyses only
Participants	
Interventions	
Outcomes	
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Kirby 1997b

Methods	RCT, Abstinence-only vs. Usual Care. Unit of randomization: School. Schools were randomized to receive adult-led PSI or usual care Method of randomization: Not reported. Allocation concealment: B (Unclear). Assessment by self-report surveys in English and Spanish, completed in school Assessment times: Baseline and 17m. Complete case analysis. Attrition: 26.3% at 17m. No significant differences among dropouts from intervention and control groups in all 3 designs; approx 1% more were lost from the intervention than from the control group in all 3 designs Baseline differences: Intervention youth were more likely to be Latino, received slightly higher grades in school, were less likely to speak only English in the home, had mothers with less education, and were more likely to have ever had sex. Intervention youth who were sexually active had slightly more sexual partners than controls. Controlled for baseline values in analyses
Participants	California (USA) N = 5244 participants at baseline. At 17-month follow-up, clusters were 31 schools Mean age: 12.7 yrs. Ethnicity: 6.2% American Indian, 11.6% Asian/Pacific Islander, 10.2% African-American, 21.3% Hispanic, 49.1% White, and 7.9% Other (percentages do not add to 100%) SES: Not reported. 42.0% Male. Notes on recruitment: Contractors, including schools and community-based organizations, applied to the California Office of Family Planning for funding to implement PSI and ENABL. They were selected because they served communities with high teen birthrates and diversity, and they demonstrated the ability to deliver PSI
Interventions	1: "Postponing Sexual Involvement (PSI)" as part of "Education Now and Babies Later (ENABL)" initiative. PSI was implemented in addition to whatever standard sexuality curriculum the individual school offered. Students received instruction in human sexuality, then received 5-session curriculum in class delivered by adults. Concepts included risks of early sex, reasons to have sex or wait, resisting social pressures,

Kirby 1997b (*Continued*)

	limit-setting, assertive responses to resist sexual pressures. Activities included class discussions, group activities, videos, slides, and roleplays. Theoretical basis: Not reported, but original program designed using social inoculation theory. Setting: school classes (courses included science, health, physical education, and others) Exposure: 5 classes, 45-60 mins long. 4 sessions were designed for delivery over 4 consecutive weeks, with the 5th session delivered between one week and one month later. Staff: Professional educators or college interns. Approximately 40% of adult leaders were health educators, 23% were classroom teachers, 23% were student interns, and the others were nurses, youth workers, or volunteers 2: Usual care - standard sexuality curriculum offered by school (varied). Generally was not comparable to PSI
Outcomes	STD diagnosis among all participants Pregnancy incidence among all participants Mean change in number of acts of intercourse in last 3m among participants reporting sexual experience at baseline or follow-up Mean change in number of acts of intercourse in last 12m among participants reporting sexual experience at baseline or follow-up Mean change in number of partners ever among participants reporting sexual experience at baseline or follow-up Condom use at last act of sexual intercourse among participants reporting sexual experience at baseline or follow-up Ever had sex among participants who were sexually inexperienced at baseline
Notes	Funding: Main funding from the California Office of Family Planning. Additional support from the Cowell Foundation, the Packard Foundation, and the Stuart Foundations Part of 3-design study. In other designs, youth were randomized by classroom and on an individual basis Extensive reporting on implementation challenges and participant satisfaction. PSI was delivered in the context of ENABL, a large statewide media and school-based campaign intended to raise awareness of teen pregnancy, encourage involvement in ENABL initiatives, and encourage a more supportive environment for youth sexual abstinence. Program contractors often established "community coalitions" and conducted community awareness activities

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Kirby 1997c

Methods	RCT, Abstinence-only vs. Usual Care. Unit of randomization: Individual. Individual youth were assigned to adult-led PSI or nonintervention control Method of randomization: Not reported. Allocation concealment: B (Unclear). Assessment by self-report surveys in English and Spanish. Assessment times: Baseline and 17m. Complete case analysis. Attrition: 45.0% at 17m. No significant differences among dropouts from intervention and control groups in all 3 designs; approx 1% more were lost from the intervention than from the control group in all 3 designs No significant baseline differences. Controlled for baseline values in analyses
Participants	California (USA) N = 704 participants at baseline. At 17-month follow-up, youth were in 17 community agencies Mean age: 13.5 yrs. Ethnicity: 1.6% American Indian, 49.4% Asian/Pacific Islander, 2.3% African-American, 19.9% Hispanic, 7.8% White, 3.6% Other (percentages do not add to 100%) 45.4% Male. Notes on recruitment: Contractors, including schools and community-based organizations, applied to the California Office of Family Planning for funding to implement PSI and ENABL. They were selected because they served communities with high teen birthrates and diversity, and they demonstrated the ability to deliver PSI. Participants were recruited through contractors' usual channels
Interventions	1: "Postponing Sexual Involvement (PSI)" as part of "Education Now and Babies Later (ENABL)" initiative. PSI was implemented in addition to whatever standard sexuality curriculum the individual school offered. Students received instruction in human sexuality, then received 5-session curriculum in class delivered by adults. Concepts included risks of early sex, reasons to have sex or wait, resisting social pressures, limit-setting, assertive responses to resist sexual pressures. Activities included class discussions, group activities, videos, slides, and roleplays. Theoretical basis: Not reported, but original program designed using social inoculation theory. Setting: community-based organizations (churches, social service agencies, other community settings Exposure: 5 classes, 45-60 mins long. 4 sessions were designed for delivery over 4 consecutive weeks, with the 5th session delivered between one week and one month later. Staff: Professional educators or college interns. Approximately 40% of adult leaders were health educators, 23% were classroom teachers, 23% were student interns, and the others were nurses, youth workers, or volunteers 2: Usual care - standard sexuality curriculum offered by community agency (varied). Generally was not comparable to PSI
Outcomes	STD diagnosis among all participants Pregnancy incidence among all participants Mean change in number of acts of intercourse in last 3m among participants reporting sexual experience at baseline or follow-up Mean change in number of acts of intercourse in last 12m among participants reporting sexual experience at baseline or follow-up Mean change in number of partners ever among participants reporting sexual experience at baseline or follow-up Condom use at last act of sexual intercourse among participants reporting sexual experience at baseline

Kirby 1997c (*Continued*)

	or follow-up Ever had sex among participants who were sexually inexperienced at baseline
Notes	Funding: Main funding from the California Office of Family Planning. Additional support from the Cowell Foundation, the Packard Foundation, and the Stuart Foundations Part of 3-design study. In other designs, youth were randomized by classroom and by school Extensive reporting on implementation challenges and participant satisfaction. PSI was delivered in the context of ENABL, a large statewide media and school-based campaign intended to raise awareness of teen pregnancy, encourage involvement in ENABL initiatives, and encourage a more supportive environment for youth sexual abstinence. Program contractors often established "community coalitions" and conducted community awareness activities

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Miller 1993

Methods	RCT, Abstinence-only enhanced (with newsletters) vs. Abstinence-only nonenhanced (no newsletters) vs. Nonintervention Unit of randomization: Family. There appeared to be only one "target adolescent" per family Method of randomization: Not reported. Distribution to treatment arms was unequal: 126 families were assigned to the videos plus newsletters, 132 families were assigned to receive the videos only, and 290 families were assigned to control group Allocation concealment: B (Unclear). Assessment by self-report surveys administered by home visitor. Surveys were administered separately to parents and teens Assessment times: Baseline, 3m, and 12m. Complete case analysis. Attrition 8.2%; attrition analyses not reported. No explicit statement of baseline differences.
Participants	Northern Utah (USA) N = 548 families at baseline. Analyses represent 503 participants Mean age: 12.9 yrs. Ethnicity: Not reported for adolescents, but reported for parents. Mothers: 93% Caucasian, 85% Mormon. Fathers: 97% Caucasian, 88% Mormon SES: Not reported. % Male: Not reported. Notes on recruitment: 6,000 families in two semi-rural and two urban school districts were invited by letters to participate. These consisted of all families with a 7th or 8th grade child in school
Interventions	1: "Facts and Feelings" enhanced. Video-based intervention. Concepts included puberty, sexual values, facts about sex, meanings of sex in relationships, consequences of sex, and assertiveness skills. Activities included video viewing, family discussion, and receipt of newsletters with discussion suggestions. Frequency of newsletter receipt was unclear. A project staff member delivered and explained the materials at the families'

Miller 1993 (*Continued*)

	homes, made follow-up calls every 2 wks for 3m, and collected the materials after 3m. Theoretical basis: Not reported. Setting: Family home. Exposure: 6 sessions, each with a 15-20 min video segment. Staff: A project staff member delivered tapes and made follow-up phone calls 2: "Facts and Feelings" without newsletters: As above, but without mailed newsletters 3: Nonintervention control.
Outcomes	Sexual initiation, as evaluated within a sexual risk behavior index. the index included behaviors ranging from handholding to sexual intercourse
Notes	Funding: US Department of Health and Human Services, Office of Adolescent Pregnancy Programs the two active intervention groups reportedly did not differ in self-reported video exposure Unclear to what extent youth had privacy from parents during home-based assessments Homogenous sample by ethnicity and religion. Disparity between trial arm sizes.

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Trenholm 2007a

Methods	RCT, Abstinence-only vs. Usual care Unit of randomization: Individual Method of randomization: Random number generator Allocation concealment: B (Unclear). Assessment by written self-report survey administered by trained data collectors. No personal information appeared on the survey. Surveys were sometimes administered by phone Assessment times: Baseline, 62.5m. Complete case analysis (secondary analyses also conducted by treatment exposure) Attrition: 19%, higher among controls. Group equivalence: Baseline differences in percentage "other" ethnicity; controlled in analyses
Participants	Powhatan, VA (USA) N = 551at baseline. Mean age: 13.3 yrs. Ethnicity: 11% African-American, 3% Hispanic, 83% white. SES: middle/working class. 49% Male Notes on recruitment: Program was part of a middle-school and high-school curriculum
Interventions	1: My Choice, My Future! In-school curriculum composed of three separate programs: Reasonable Reasons to wait (focused on character, marriage, STIs, relationships, parenthood, dating, and human development) , the Art of Loving Well (focused on love, romance, commitment, and marriage and included short stories and fairy tales), and WAIT Training Workshop (focused on love, sexuality, media, abstinence, intimacy,

Trenholm 2007a (*Continued*)

	parental communication, HIV, and the consequences of sex). Theoretical basis: Not reported. Setting: Middle and high schools Exposure: 52 class sessions over 3 years. Staff: Appeared to be led by adults. 2: Usual care defined by schools. Included nine-week health and physical education class that did not directly discuss abstinence or STIs, along with an additional class that covered abstinence, but not STIs or contraceptive use
Outcomes	STD diagnosis among participants who have ever had sex Pregnancy incidence among participants who have ever had sex Unprotected sex in the past year (i.e., sex without consistent condom use) among all participants Sex in the past year among all participants Number of partners ever among all participants Condom use at first sex among participants who have ever had sex Ever had sex among all participants
Notes	Funding: US Department of Health and Human Services, commissioned by the US Congress Follow-up intervals varied for different participants; calculated here by subtracting the mean baseline participant age from the mean age at follow-up

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Trenholm 2007b

Methods	RCT, Abstinence-only vs. Usual care Unit of randomization: Individual Method of randomization: Random number generator Allocation concealment: B (Unclear). Assessment by written self-report survey administered by trained data collectors. No personal information appeared on the survey. Surveys were sometimes administered by phone Assessment times: Baseline, 65m. Complete case analysis (secondary analyses also conducted by treatment exposure) Attrition: 20%, higher among controls. Group equivalence: Baseline differences in "chance will have sex next year" and "chance will have sex before end of high school"; controlled in analyses
Participants	Miami, FL (USA) N = 597 at baseline. Mean age: 12.8 yrs. Ethnicity: 63% African-American, 3% Hispanic, 3% white. SES: low. 0% Male Notes on recruitment: Program was an optional elective in a middle-school curriculum

Trenholm 2007b (*Continued*)

Interventions	1: ReCapturing the Vision. Daily in-school elective course for girls that included two separate programs: ReCapturing the Vision (focused on self-esteem, goal-setting, values, careers conflict resolution, and the transition to adulthood), and Vessels of Honor (focused on honor, abstinence, refusing sex, consequences of sex, peer pressure, relationships, puberty, sexual abuse, date rape, choosing a mate, and marriage). Theoretical basis: Not reported. Setting: Middle school classrooms. Exposure: Approximately 180 sessions over 1 year. Staff: Appeared to be led by adults. 2: Usual care defined by schools. Included week-long unit on human growth and development, STIs, abstinence, and drug and alcohol prevention
Outcomes	STD diagnosis among participants who have ever had sex Pregnancy incidence among participants who have ever had sex Unprotected sex in the past year (i.e., sex without consistent condom use) among all participants Sex in the past year among all participants Number of partners ever among all participants Condom use at first sex among participants who have ever had sex Ever had sex among all participants
Notes	Funding: US Department of Health and Human Services, commissioned by the US Congress Program attendance was low: 35% of intervention-group participants did not participate in any program activities Follow-up intervals varied for different participants; calculated here by subtracting the mean baseline participant age from the mean age at follow-up

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Trenholm 2007c

Methods	RCT, Abstinence-only vs. Usual care Unit of randomization: Individual Method of randomization: Random number generator Allocation concealment: B (Unclear). Assessment by written self-report survey administered by trained data collectors. No personal information appeared on the survey. Surveys were sometimes administered by phone Assessment times: Baseline, 62.5m. Complete case analysis (secondary analyses also conducted by treatment exposure) Attrition: 18%, higher among controls. Group equivalence: Baseline differences in age, group differences in timing of final follow-up interview; controlled in analyses
Participants	Milwaukee, WI (USA) N = 504 at baseline. Mean age: 10.3 yrs.

	Ethnicity: 77% African-American, 8% Hispanic, 2% white. SES: low 38% Male Notes on recruitment: Program was an optional after-school activity in elementary and middle schools
Interventions	1: Families United to Prevent Teen Pregnancy. Curriculum met daily over 4 years for approximately 2.5 hours each day. Focused on self-esteem, values, goals, decision-making, risky behaviors, communication, relationships, sexuality, puberty, STIs, and social skills. Theoretical basis: Not reported. Setting: Elementary and middle schools. Exposure: Approximately 720 sessions over 4 years. Staff: Appeared to be led by adults. 2: Usual care defined by schools. Included mandatory family life education for all grades, which incorporated units on abstinence and contraceptive use beginning in grade 5
Outcomes	STD diagnosis among participants who have ever had sex Pregnancy incidence among participants who have ever had sex Unprotected sex in the past year (i.e., sex without consistent condom use) among all participants Sex in the past year among all participants Number of partners ever among all participants Condom use at first sex among participants who have ever had sex Ever had sex among all participants
Notes	Funding: US Department of Health and Human Services, commissioned by the US Congress Program attendance was low: 43% of intervention-group youth did not participate in any program activities. of the 57% that participated to any extent, 55% participated in fewer than half of program activities the first year. the average program youth who attended any program services received approximately 146 hours of program services in the first year Follow-up intervals varied for different participants; calculated here by subtracting the mean baseline participant age from the mean age at follow-up

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Trenholm 2007d

Methods	RCT, Abstinence-only vs. Usual care Unit of randomization: Individual Method of randomization: Random number generator Allocation concealment: B (Unclear). Assessment by written self-report survey administered by trained data collectors. No personal information appeared on the survey. Surveys were sometimes administered by phone Assessment times: Baseline, 59m. Complete case analysis (secondary analyses also conducted by treatment exposure) Attrition: 16%, higher among experimental group. Group equivalence: Baseline differences in knowledge of STIs; controlled in analyses
Participants	Clarksdale, MS (USA) N = 849 at baseline. Mean age: 10.7 yrs. Ethnicity: 87% African-American, 8% Hispanic, 0% white. SES: low 48% Male Notes on recruitment: Program was a mandatory part of the curriculum in elementary schools
Interventions	1: Teens in Control. Mandatory in-school curriculum composed of two separate programs: Postponing sexual involvement (as in Kirby 1997a, Kirby 1997b, Kirby 1997c; focused on risks of sex, STIs, social pressures, peer pressure, assertiveness, and skills), and Sex Can Wait (as in Goldfarb 1999; focused on self-esteem, puberty, communication, sex, goal setting, life planning, and family values). Theoretical basis: Not reported. Setting: Elementary school classrooms. Exposure: Approximately 72 sessions over 2 years. Staff: Appeared to be led by adults. 2: Usual care defined by schools. Included limited health, family-life, and sex education curricula for youth in middle school
Outcomes	STD diagnosis among participants who have ever had sex Pregnancy incidence among participants who have ever had sex Unprotected sex in the past year (i.e., sex without consistent condom use) among all participants Sex in the past year among all participants Number of partners ever among all participants Condom use at first sex among participants who have ever had sex Ever had sex among all participants
Notes	Funding: US Department of Health and Human Services, commissioned by the US Congress. Follow-up intervals varied for different participants; calculated here by subtracting the mean baseline participant age from the mean age at follow-up

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abel 2007	Nonrandom assignment of participants. Location: Florida n=not reported Mean age: Grades 5-9 Follow-up: Immediate post-intervention Intervention: Abstinence only vs. Unclear control group. Family Action Model for Empowerment, focused on 12-session after-school program and 9-week parent education group. Findings: No behavioral outcomes presented.
Barnett 2003	Nonrandom assignment of participants. Location: Northwest Missouri n=56 Mean age: 8th grade (approx 13 yrs) Follow-up: 4 months Intervention: Abstinence-only vs. Nonintervention (delayed intervention control). Life's Walk Program - 3-week in-school program including infant simulators. Findings: No statistically significant differences on behavior scale, which included behaviors ranging from "kissing and hugging" to "touching beneath the waist under clothing." No questions about specific sexual behaviors
Bearman 2001	Nonrandom assignment of participants; cross-sectional survey comparison. Location: National US sample drawn from schools n=90,000 students completed in-school surveys. 20,745 of these students completed in-home surveys for wave 1, and 14,787 of the in-home students were resurveyed at wave 2. Age range: Grades 7-12 at Wave 1. Wave 2 occurred 1 year later. Follow-up: 1 year Intervention: Abstinence-only (reported virginity pledge) vs. No reported virginity pledge Findings: Significantly protective protects of pledgers were observed for sexual initiation in analyses controlling for personal characteristics, religiosity, and other protective factors. Pledgers' relative risk of sexual initiation was 34% lower than non-pledgers' relative risk
Bekenstein 2007	Nonrandom assignment of participants. Location: Milwaukee, WI n=360 Mean age: Grades 6-12 Follow-up: Immediate post-intervention Intervention: Abstinence-only vs. Unclear comparison group. Teens Taking Charge, includes school- and community-based abstinence education groups, youth leadership, peer education, and family/community involvement. Findings: No behavioral outcomes presented.
Bersamin 2005	Nonrandom assignment of participants; cross-sectional survey comparison. Location: National US sample drawn from schools n=870 Age range: 12-16 yrs Follow-up: First 3 waves of 5-wave longitudinal survey design. Used data from the National Longitudinal

(*Continued*)

	Survey of Adolescent Health (like Bruckner 2005). Intervention: Abstinence only (formal or informal virginity pledge) vs. No reported virginity pledge. Findings: Students who had made private pledges to delay sex were less likely to initiate oral and vaginal sex than those who had not. Making a formal pledge did not affect sexual behavior
Borawski 2001	Nonrandom assignment of participants. Full text unavailable at the time of this review; data were abstracted from Rector 2002 and Kirby 2002. Location: Cleveland, OH n=>800 Target age range: 12-13 years Follow-up: 14 weeks post-intervention Intervention: Abstinence-only vs. Control (unclear what treatment the control group received, if any). Operation Keepsake. Findings: No significant program effects on sexual initiation or returning to abstinence. Comparisons for both outcomes non-significantly favored program participants; analyses did not control for baseline differences
Borawski 2005	Nonrandom assignment of participants. Location: Cuyahoga County, OH n=3,017 at baseline; 2,069 analyzed Age range: Grades 7-8 (approx 11-14 yrs). Mean age approx 13 yrs. of participants analyzed after attrition, 37.6% were younger than 13 yrs. Follow-up: 5 months (mean 129 days) after intervention Intervention: Abstinence-only vs. Nonintervention (delayed intervention control group). Operation Keepsake - 5-session in-school curriculum focusing on benefits of abstinence, character development, future life goals, and resistance skills. Curriculum emphasizes limitations of condoms and contraception. Findings: Significantly favored program participants for STD/HIV knowledge ($p < 0.001$). No significant program effects for recent sex (i.e., sex since baseline) among all participants (OR=0.85 [95% CI=0.62, 1.15]). Subgroup analyses showed no significant effects among participants who reported no sexual experience at baseline (n=1462, OR=0.83 [0.52, 1.33]), and among participants who did report sexual experience at baseline (n=439, OR=0.87 [0.58, 1.31]). Among participants who reported sex between pre-test and post-test (n=311), intervention participants reported significantly lower frequency of vaginal sex (beta=-1.74, SE=0.83, $p < 0.05$). They were also less likely to report multiple episodes of sex (6 or more vs. 5 or less), (OR=0.47 [0.26, 0.84], $p < 0.05$) and to report two or more sexual partners (OR=0.50 [0.30, 0.83], $p < 0.01$). No significant effect on consistent condom use (OR=1.19 [0.71, 1.99])
Bruckner 2005	Nonrandom assignment of participants; cross-sectional survey comparison. Location: National US sample drawn from schools n=15,170 surveyed; 11,471 had valid data on STD status Age: ages 11-19 at first survey administration (1995), age 18-24 yrs at third wave (2001-2002). Follow-up: 3 waves of longitudinal survey design, 6.5 years elapsed between Wave 1 and Wave 3. Used data from the National Longitudinal Survey of Adolescent Health (like Bersamin 2005). Intervention: Abstinence-only (reported virginity pledge) vs. No reported virginity pledge. Findings: 88% of pledgers have sex before marriage, compared to 99% of non-pledgers. Used bio-marker data to test for STDs. No significant differences in STD infection rates between pledgers and non-pledgers. Among married respondents, STD rates also did not differ by pledge status. Pledgers married significantly earlier than non-pledgers. Pledgers were significantly less likely to have used a condom at first sex, but no significant differences were observed in condom use at most recent sex or condom use over the past 12 months before Wave 3 surveys. Pledgers reported sexual initiation at a later age, fewer partners, and lower numbers of non-monogamous partners than non-pledgers. Significantly more pledgers than non-pledgers reported

(*Continued*)

	having had oral sex with one or more partners but no vaginal sex. Significantly more male pledgers than male non-pledgers reported having had anal sex but no vaginal sex. Respondents reported condom use for first oral sex for only 4% of relationships involving oral sex. Respondents reported condom use for first anal sex for only 30% of relationships involving anal sex. Hypothesized that the unexpected higher STD acquisition among pledgers was due to the combination of low condom use for oral and anal sex and pledgers' higher likelihood of substituting these acts for vaginal sex
Carter 2004	Lack of control group; comparisons made to surveys. Location: Kansas n=1,241 Age range: 9-19 yrs Follow-up: Immediate post-intervention Intervention: Abstinence-only programs implemented by 6 different contractors in Kansas (Kansas Abstinence Education Program). Programs varied among six contractors. Findings: Pre-post comparisons show no significant difference in the percentage of participants reporting that they are currently sexually active. Statewide survey data show trends towards decreased pregnancies, but no evidence of decline in statewide STI incidence; specific comparisons between program participant data and statewide survey data were not conducted
Christopher 1990	Nonrandom assignment of participants. Location: Phoenix, AZ n=320 Mean age: 12.8 yrs. Follow-up: Immediate post-intervention. (6 weeks post baseline) Intervention: Abstinence-only vs. Usual care. Success Express - 6-session in-school program supplementing regular health curriculum, teach behaviors, attitudes, and skills. Supported by Adolescent Family Life Act (1981). Findings: Intervention group significantly increased mean sexual interaction levels (specific increases in touching female breasts, touching female genitals, and genital-genital contact), while controls did not. the finding was repeated when non-virgins were excluded from the analysis
Denny 1999	Nonrandom assignment of participants. Location: Arkansas. n=2,335 Ages: Divided into upper elementary, middle school, and high school students. Follow-up: Immediate post-intervention. Intervention: Abstinence-only vs. Usual Care. Sex Can Wait - 5-wk in-school curriculum compliant with federal definition of abstinence education. Separate curriculum for each age group. Findings: Behavioral outcomes limited to middle school and high school participants. No significant group differences for middle school participants reporting sex in the last 30 days. No significant group differences for high school participants reporting sex in the last 30 days
Denny 2002	Nonrandom assignment of participants. Location: Arkansas (exact location was unclear). n=1,421 Ages: Divided into upper elementary, middle school, and high school students. Follow-up: 1 to 2 months, 18 months Intervention: Abstinence-only vs. Usual Care. Sex Can Wait - 5-wk in-school curriculum compliant with federal definition of abstinence education. Separate curriculum for each age group.

(*Continued*)

	Findings at 1-2 months: Measured incidence of sex ever and incidence of sex in the last 30 days. No significant behavioral differences in behaviors of upper elementary students. No significant behavioral differences in behaviors for middle school students. for high school students, when analyses controlled for virgin status and behavioral intent, significantly fewer students reported sex ever and sex in the past 30 days at follow-up. Findings at 18 months: for upper elementary participants, intervention group had significantly higher knowledge scores and were less likely to report sex in the past month. No difference in ever had sex. for middle school participants, intervention group had significantly higher knowledge scores and were less likely to have ever had sex, and to have had sex in the last month. for high school participants, intervention group had significantly higher knowledge scores, but no differences in sex ever or sex in the last month Analyses were ANCOVAs for knowledge and logistic regression for behavior
Denny 2003	Nonrandom assignment of participants. Location: 50 Arkansas school districts n > 6,000 Ages: Divided into upper elementary, middle school, and high school students. Follow-up: 18 months Intervention: Abstinence-only vs. Usual Care. Sex Can Wait - 5wk in-school curriculum compliant with federal definition of abstinence education. Separate curriculum for each age group. Findings: Significantly fewer middle school participants reported sex ever or recent sex than middle school controls. Significantly fewer upper elementary school participants reported recent sex than elementary school controls; not significant for sex ever for this group. No significant effect on sex ever or recent sex in high school population
DiFiore 2007	Nonrandom assignment of participants. Location: New Jersey n=1516 enrolled; 1444 completed immediate post-intervention assessment; 483 completed 10-week follow-up survey. Mean age: unclear Follow-up: 10 weeks Intervention: Abstinence only vs. Usual care. the Choice Game: focused on sex, peer pressure, drugs/alcohol, pregnancy, and STDs. Findings: No behavioral outcomes presented.
Doniger 2001	Lack of control group; comparisons to survey data. Location: Monroe County, NY n=1,578 middle school students and 1,737 high school students at Wave 3. Random sampling of middle and high school students; not designed to track individuals. Age range: grades 7-12 (12-18 yrs). Follow-up: Cross-sectional time series design, with 3 waves of surveys. Final wave took place after the third program year. Intervention: Abstinence-only. “Not Me, Not Now” - Multi-component intervention including radio ads, posters, the Postponing Sexual Involvement in-school curriculum, parent packet, website, educator forums, and community events. Findings: No behavioral outcomes reported for middle school students. for random samples of high school students, age of sexual initiation significantly declined between Wave 1 and Wave 3. Frequency of self-reported sexual intercourse by age 15 significantly declined; however, this was not significant for frequency of self-reported sexual intercourse by age 17. Adolescent pregnancy rates in Monroe County from 1993 to 1996 were compared to those in New York State, Upstate New York, and two large New York counties. Using chi-square test, four of the five areas had

(*Continued*)

	statistically significant downward trend in adolescent pregnancy. the slope of the regression line for Monroe County was significantly steeper than the slope of any other regression line
Goodson 2004	Lack of control group; comparisons to survey data. Location: Texas n=451 middle school students, 277 high school students. Mean age: 13.29 yrs for middle school, 15.5 yrs for high school. Follow-up: Ranged between several weeks and several months. Intervention: Abstinence-only vs. Usual Care. Various abstinence-only interventions delivered by 5 organizations were grouped for evaluation. Findings: Many pre-post outcomes were not tested for statistical significance. When controlled for age and dosage, middle school participants reported increased frequencies at post-test for sex ever and ever having been involved in "petting." High school students reported increased frequencies of sex ever and oral sex ever at post-test. the middle school program sample had lower rates of sex ever by age 15 than a national sample of youth. the high school sample had lower rates of sex ever by age 18 than a national sample of youth. Exceptions: Females in the high school sample reported higher rates of sex ever by age 16 than a national sample of youth. Males in the high school sample reported higher rates of sex ever by age 17 than a national sample of youth
Halpin 2007	Not a prospective RCT (all participants received the intervention, but were randomly assigned to complete surveys at pre-test or post-test; they were not randomly allocated to conditions). Location: Rural area in Alabama n=1425 Mean age: 7th and 8th grade Follow-up: Immediate post-intervention Intervention: Abstinence-only (surveyed after intervention) vs. Abstinence-only (surveyed before intervention). Choosing the Best Life, includes 5 adult-led classroom sessions. Findings: No behavioral outcomes presented.
Hubbard 1998	Nonrandom assignment of participants. Location: "Southern Rural State" - probably Arkansas due to funding from Arkansas State Dept of Education. n=212 Age range: Grades 9-10. (57% were aged 15-16yrs) Follow-up: 18m Intervention: Abstinence-plus vs. Abstinence-only and state-approved curricula. All programs were one semester long; no details on what the abstinence-only curricula entailed or what proportion of the control group actually received abstinence-only programs. Findings: Among baseline virgins, significantly fewer students in the abstinence-plus group became sexually active at posttest than abstinence-only and state-curriculum controls. Among participants who became sexually active post-intervention, significantly more abstinence-plus participants reported using STD/HIV and pregnancy prevention than abstinence-only and state-curriculum controls
Janken 2005	Nonrandom assignment of participants. Location: Location: North Carolina n=245 Mean age: 6th-8th grade Follow-up: 3 and 12 months Intervention: Abstinence-only vs. Unclear control group. Teaching Responsible Actions in Life: in-class curriculum about risk behaviors, with additional tutoring, summer activity placement, support groups, and

(*Continued*)

	peer education. Findings: No behavioral outcomes presented.
Jorgensen 1993	Focus on pregnancy prevention; HIV or STD prevention was not a stated program goal. Location: Delaware, Mississippi (and Ohio at 6-week follow-up) n=91 Mean age: 14.4yrs at baseline. Follow-up: 6 weeks and 6 months Intervention: Abstinence-only vs. Usual Care. "Project Taking Charge" - In-school instruction and parent-youth sessions targeting abstinence and vocational planning. Findings: No outcomes of interest at 6-wk follow-up. At 6-month follow-up, marginally significant effect on initiation of sexual activity among baseline virgins, favoring program (p=.051, n=50)
Lage 2005	No outcomes of interest. Location: Appears to be Florida n=92 Mean age: 9-14 years Follow-up: 6 months Intervention: Abstinence-only vs. Usual care. Multimedia Approach to Pregnancy Prevention (MAPP) for chronically ill youth. Focused on learning about disease, sexual development, pregnancy, decision making, relationships, drugs/alcohol, and communication. Findings: No behavioral outcomes presented.
LeCroy 2003	Lack of control group; comparisons to survey data. Location: Arizona n=105,090 children, preteens, teens, high-risk adults, and parents over 5 years of evaluation Age: 94% of sample were teens in grades 7-12 (aged 12-18yrs). 3% of sample were adults/parents, 3% of sample were children younger than 7th grade. Follow-up: Cross-sectional design with surveys over 5 years, not designed to track individual participants. Intervention: Abstinence-only programs delivered in schools, after-school programs, health centers, religious centers, and via media campaigns. "Arizona Abstinence Education Program." Findings: One-time follow-up survey of "select Year 4 participants" shows that 95% of teenage baseline virgins remained abstinent, while 52% of teenage baseline non-virgins remained abstinent. (Unclear how this sample was selected.) In one year, live birth rates among female program participants by age cohort were "lower than comparable state rates." Non-marital birth rate for female program participants in 2001 (collected via the appearance of participants' names on birth certificates) were lower than rate for female non-participants matched for age. Behavioral outcomes for high-risk adults were not recorded
Lerner 2004	Lack of control group; comparisons to survey data. Full text unavailable at the time of this review; data abstracted from the executive summary. Location: Washington, DC n=2,730 Follow-up: Survey data collected at the end of each program year. Intervention: Abstinence-only vs. age-matched groups in a national survey (using Youth Risk Behavior Survey data). Best Friends Program and Diamond Girls Leadership Program - one-year curriculum emphasizing decision-making, avoiding alcohol and drugs, and avoiding sex. Best Friends is the middle school curriculum, and Diamond Girls is the high school curriculum. Findings: for the Best Friends comparison, intervention group significantly more likely to report not having had sex than the national sample (OR=6.48, $p < 0.001$). This finding was repeated for the Diamond Girls

(*Continued*)

	comparison (OR=119.58, p < 0.001)
Lieberman 2005	Nonrandom assignment of participants. Location: New York, NY n=445 Mean age: 5th-8th grade Follow-up: Immediate post-intervention Intervention: Abstinence-only vs. Usual care. NiteStar StarLo: Theater-based sessions focusing on puberty and family issues in accordance with federal abstinence-only requirements. Findings: No behavioral outcomes presented.
Lin 2007	Nonrandom assignment of participants Location: Montgomery County, AL n=324 Mean age: 7th and 10th grade Follow-up: Immediate post-intervention Intervention: Abstinence-only vs. No treatment. Findings: No behavioral outcomes presented.
Minnesota 2002	Nonrandom assignment of participants. Location: Minnesota n=Unclear. One set of tables indicates that 468 students surveyed; text indicates that Age range: Grade 9 (14-15yrs) and Grade 12 (17-18yrs) Follow-up time: Cross-sectional survey approximately 1 year after programs began. Intervention: State distributed contracts for abstinence-only funding. Cross-sectional survey comparing students in counties receiving Abstinence-only (no implementation) vs. Abstinence-only (weak/moderate implementation) vs. Abstinence-only (strong implementation). Program designs varied by contractor. Interventions were aimed at youth with the target age 10-14; 12th-graders were actually too old to participate in the curricula. Findings: In a cross-tabulation of the frequency of self-reported sexual activity (i.e., none, 1-2 times, 3 or more times - unclear what the recall period was) with the degree of program implementation by county (i.e., none, weak, strong), the relationship was significant for 12th grade students. Means suggest that students in counties with strong implementation were less likely to report any sex than participants in weak-implementation counties, who were less likely to report any sex than participants in no-implementation counties. This was true for males and females. Analyses were non-significant for 9th-grade students (i.e., the age group that was intended to participate in the curricula)
Mitchell-DiCenso 97	Focus on pregnancy prevention; HIV or STD prevention was not a stated program goal. Location: Hamilton, Ontario (Canada) n=3,975 students in 21 schools Mean age: 12.6 years at baseline. Follow-up: Yearly follow-ups for 4 years post-intervention. Intervention: Abstinence-only vs. Usual Care. "McMaster Teen Program" 10-session school-based intervention delivered by public health nurses, concepts including peer pressure, adolescent development, gender roles, stages of intimacy, consequences of sex, childbearing, teen pregnancy, communication, problem-solving, decision-making, and relationships. No description of contraceptive methods or condoms. Activities included films, discussions, roleplays, Q&A sessions. Findings: No significant differences by group for male or female participants in survival curves for time to first sex. No significant differences by group in survival curves for time to first pregnancy for females

(*Continued*)

Nicholson 1992	Nonrandom assignment of participants. Location: 4 sites in Dallas, TX; Memphis, TN; Omaha, NE; and Wilmington, DE n=99 participants who attended only the abstinence-only program components Mean age: 12-15 when study began Follow-up: 2 years Intervention: Growing Together (abstinence-only program, 15 participants) vs. Will Power/Won't Power (abstinence-only program, 84 participants) vs. nonintervention (controls, 106 participants). Interventions were two of the four community-based programs offered by Girls Incorporated. Growing Together had 1 parent session and four 2-hour parent-daughter sessions, Will Power/Won't Power had six 2-hour sessions. the other two program offerings contained contraception instruction. Findings: Pregnancy rates 13.3% in participants only attending Growing Together; 3.6% in participants only attending Will Power/Won't Power; 12.3% in participants attending no programs
O'Donnell 2005	No outcome data of interest. Location: 7 New York City schools n=846 youths and their families Age range: 83% of participants were aged 10-11 Follow-up: 9 months after baseline Intervention: Saving sex for later vs. No treatment. A series of 3 CDs were mailed to intervention family homes over 6 months. CDs are based on the theory of planned behavior and encourage parents to talk about sexual abstinence with their children. Focus on parent-child conversations about puberty, relationships, and peer pressures. Findings: A "risk scale" was used to assess 10 behaviors including "hung out with opposite-sex peers disapproved by a parent, had a girlfriend or boyfriend, and kissed and hugged for a long time." Scale ranged from 0-4. At 9-month follow-up, "youths' intervention status was significantly related to more family rules, more family support, and fewer risk behaviors." from regression analysis, OR=0.36 (95% confidence interval 0. 29, 0.43), p<0.001 for risk behavior index, favoring the intervention group
Olsen 1991	Nonrandom assignment of participants. Location: 3 school districts in Utah n=Not reported. All 7th- and 10th-grade students in three districts were surveyed, some 11th- and 12th-grade participants in one site. Ages: 7th grade and 10th grade (approx 12-13 yrs and 15-16 yrs) Follow-up: Immediate post-intervention Interventions: Abstinence-only. Each group received 1 of 3 school-based programs - Sex Respect (10 sessions) , Values and Choices (15 sessions), Teen-Aid (15 sessions) Findings: No behavioral outcomes presented.
Olsen 1992	Nonrandom assignment of participants. Appears to duplicate Olsen 1991, but unclear. Location: 3 school districts in Utah n=Not reported. All 7th- and 10th-grade students in three districts were surveyed, some 11th- and 12th-grade participants in one site. Ages: 7th grade and 10th grade (approx 12-13 yrs and 15-16 yrs) Follow-up: Immediate post-intervention Interventions: Abstinence-only. Each group received 1 of 3 school-based programs - Sex Respect (10 sessions) , Values and Choices (15 sessions), Teen-Aids (15 sessions) Findings: No behavioral outcomes presented.

(*Continued*)

Pallone 2007	No outcome data of interest; unclear if random. Location: Indianapolis, IN n=731 Mean age: 6th grade Follow-up: Immediate post-intervention Intervention: Abstinence-only vs. No treatment. Peers Educating and Encouraging Responsible Sexuality: slightly older mentors educate participants about risks and consequences of premarital sex and value of abstinence. Findings: No behavioral outcomes presented.
Pruitt 2005	Lack of control group; comparisons to survey data. Location: Texas n=728 students Age range: 11-17yrs Follow-up time: Unclear; ranged from immediate follow-up until 1 year. Intervention: Abstinence-only vs. National Average figures. Students attended various abstinence-only programs at 29 schools. Findings: Before receiving abstinence-only education, 23% of 9th-grade girls reported having had sex (below the national average); after the programs, this figure was 28%, which was nearer to their Texas peer group. Boys reported no change at follow-up; however, during the following year, the percentage reporting sexual initiation increased from 24% to 39%
Resnick 1997	Nonrandom assignment of participants; cross-sectional survey comparison. Location: National US sample drawn from schools. n=12,118 Age range: Grades 7-12 (ages 12-18 yrs) Follow-up: Cross-sectional survey design. Intervention: Self-reported formal virginity pledge vs. No report of pledging virginity. Data drawn from National Longitudinal Study of Adolescent Health. Findings: Self-reporting a formal virginity pledge was significantly associated with later initiation of sex. This relationship was observed for a number of other characteristics as well
Roberts-Gray 2007	No outcomes of interest; no baseline assessment. Location: Texas n=approx 199 Mean age: younger than 13 Follow-up: Immediate post-intervention Intervention: Abstinence-only (year-long) vs. Abstinence-only (semester-long). LifeWorks' Adolescent Pregnancy Prevention Program, included Managing the Pressures Before Marriage and Growing Up and Making Health Choices curricula. Findings: No sexual behavioral outcomes presented.
Roosa 1990	Nonrandom assignment of participants. Location: Unclear in trial report; probably Phoenix, AZ n=528 Mean age: 13 yrs Follow-up: Immediate post-intervention. (6 weeks post baseline) Intervention: Abstinence-only vs. Usual care. Success Express - 6-session in-school program supplementing regular health curriculum, teach behaviors, attitudes, and skills. Supported by Adolescent Family Life Act

(*Continued*)

	(1981). Findings: No significant program effects on index of sexual behavior, although trend data for virgin male participants indicate non-significant increases in sexual behavior (increasing from kissing or French kissing to touching female breasts or genitals)
Rue 2005	Nonrandom assignment of participants. Location: Colorado n=807 Mean age: 9th-12th grade Follow-up: 12 months Intervention: Abstinence-only (delivered by specialist teacher) vs. Abstinence-only (delivered by regular teacher) vs. Usual care. WAIT Training, focused on self-control, relationships, boundaries, sexual intimacy, conflict resolution, and media. Findings: No difference between the three groups in initiation of sex
Saunders 2005	Nonrandom assignment of participants. Location: Iowa n=7025 for the comparison between participants in abstinence-only initiatives and participants in abstinence-plus initiatives Mean age: Participants in abstinence-only programs were in grades 6-8 (Age range approx 12-15 yrs). Unclear for abstinence-plus programs. Follow-up: Varied by program site Interventions: Abstinence-only programs granted federal abstinence-only funding vs. Abstinence-plus programs as funded by the Iowa Department of Human Services. Abstinence-only programs were delivered by 4 separate contractors. Findings: No behavioral outcomes presented. Participants in the abstinence-only (n=1087) and abstinence-plus (n=5938) programs reported the same post-test level of knowledge of STDs and AIDS (mean score=1. 59 for both groups)
Smith 2003	Nonrandom assignment of participants. Location: Pennsylvania n=All participants in 13 programs (overall programs delivered to 22,000 youth per year). These comprised 14 separate controlled designed Age: 12 yrs or older Follow-up: Program evaluations conducted over 4 years Intervention: Abstinence-only vs. Varied control groups. Pennsylvania Abstinence Education and Related Service Initiative - abstinence-only programs implemented by 13 separate contractors. Findings: 4 of 14 studies show protective effects on sexual activity (one limited to females, one limited to males). Two studies show harmful effects on sexual activity. Methodological quality of the 14 studies varied widely
Spear 1997	Nonrandom assignment of participants. Location: Arkansas n=287 Age range: Grades 5-6 (ages 10-12) Follow-up: Immediate post-intervention (5 weeks after baseline). Intervention: Abstinence-only vs. Usual care. "Sex Can Wait" - 5-week in-school curriculum with 23 lessons. Focuses on self-esteem, puberty, values, decision-making, communication, and life planning. Findings: No behavioral outcomes presented.

(*Continued*)

St Pierre 1995	Nonrandom assignment of participants. Location: National US sample n=161 surveyed at all assessment points, from 14 Boys and Girls Clubs. Analyses represent 152 participants (9 excluded due to unreported baseline virginity status). Mean age: 13.6 yrs at baseline Follow-up time: 27-month longitudinal study Intervention: Abstinence-only enhanced (program plus boosters) vs. Abstinence-only non-enhanced (program without boosters) vs. Nonintervention. Stay SMART - 12-session program aiming to prevent sexual activity and gateway drug use. Findings: Recent vaginal sex was measured on a Likert scale from 1=never to 6=in the last 24 hours; sexual frequency was measured on a Likert scale from 1=never to 5=a few times a week. the two measures were standardized and added to produce a single sexual behavior outcome. For baseline non-virgins (n=67), a significant group-by-time interaction was observed for the sexual behavior scale ($F(4, 120)=5.41$, $p < 0.001$) . Participants receiving the non-enhanced program reported significantly less sexual behavior than either the control group or the enhanced program group. For baseline virgins (n=85), no significant group-by-time interaction was observed
Sternas 2007	Nonrandom assignment of participants. Location: Newark, NJ n=643 Mean age: 6th-8th grade Follow-up: Immediate post-intervention Intervention: Abstinence-only vs. Usual care. Newark Best Friends/Best Men Adolescent Family Life Abstinence Education Program. Based on social cognitive theory, cognitive development theory. Includes group discussions, mentoring, physical fitness, community service, cultural events, role models, counseling, sexuality information, health lessons. Includes 110 hours exposure per year. Findings: No behavioral outcomes presented.
Sulak 2006	Nonrandom assignment of participants. Location: central Texas n=26125 at baseline, 24550 at immediate post-intervention Mean age: 6th-8th grades Follow-up: Immediate post-intervention (2 weeks after baseline) Intervention: Abstinence-only vs. Usual care. In-school curriculum implemented by science teachers focusing on pregnancy, STDs, abstinence, delaying sex, skill-building, character, and refusal skills. Findings: No behavioral outcomes presented.
Tanner 2007	Nonrandom assignment of participants. Location: McLennan County, Texas n=523 Mean age: 7th-8th grades during program delivery Follow-up: 5 months after program completion Intervention: Abstinence-only vs. Usual care. McLennan County Collaborative Abstinence Program, in-school curriculum receiving federal abstinence-only funding. Findings: Significantly favored program participants for willingly engaged in sex. 8th grade: 10% vs. 23%, p=0.0015 from chi-square test. 9th grade: 21% vs. 29%, p=0.05 from chi-square test

(*Continued*)

Tobin 2005	No outcomes of interest, unclear whether random. Location: Portland, OR n=482 Mean age: 7th-8th grades. Follow-up: Immediate post-intervention. Intervention: Abstinence-only (brief) vs. Abstinence-only (intensive). Brief: Promises dramatic presentation with discussion. Intensive: Same, plus the FACTS prevention curriculum in classrooms with assistance from slightly older peers. Findings: No behavioral outcomes presented.
Usera 2007	Nonrandom assignment of participants (participants were randomly selected from groups, but had already been assigned to intervention or control) Location: Rapid City, SD n=2496 Mean age: 9-14 Follow-up: Immediate post-intervention Intervention: Abstinence-only vs. Usual care. Girls Incorporated, includes curricula for Growing Together, Will Power/Won't Power, and Baby Think It Over, along with advocacy services, health care, and parental involvement. Findings: Changes in percentage of participants reporting sexual intercourse favored the intervention group (-3.9% vs. +3.4%), but significance was unclear
Weed 1989	Nonrandom assignment of participants. Report unavailable at the time of this review.
Weed 1992	Nonrandom assignment of participants. Location: Utah (Murray District, Jordan District, and Millard District) n=7,644 at baseline. Sexual behavior analyses for 12m follow-up were limited to participants who reported virginity at baseline and for whom linked pre-post data were available (n=1640). Ages: Programs were implemented in both middle schools and high schools. Middle school students were "primarily" in 7th grade (approx 12-13 yrs), and high school students were in "primarily" 10th grade (approx 15-16 yrs). for Values and Choices, participants were in 8th grade (approx 13-14 yrs). Follow-up: Posttest and 12 months. Interventions: Abstinence-only vs. Control group. It was unclear what treatment the control group received, if any. Intervention groups received one of 3 school-based programs - Sex Respect (10 sessions), Values and Choices (15 sessions), Teen-Aid (15 sessions). Findings: Sexual initiation was assumed to refer to vaginal sex. At 12m follow-up, no significant program effects were reported for sexual initiation at the high school level (n=673) or at the middle school level (n=967)
Weed 1994	Nonrandom assignment of participants. Full text unavailable at the time of this review; data abstracted from Rector 2002 and Kirby 2002. Location: Unclear n=308 Mean age: Unclear Follow-up: Unclear. Follow-up times for the treatment and comparison groups differed. Intervention: Abstinence-only vs. Control group (Unclear what treatment the control group received, if any) . Family Accountability Communicating Teen Sexuality (FACTS). Findings: According to Rector 2002, "students who participated in the program were 30% to 50% less likely

(*Continued*)

	to commence sexual activity than those who did not participate."
Weed 2001	Nonrandom assignment of participants. Full text unavailable at the time of this review; data abstracted from Rector 2002 and Kirby 2002. Location: Arkansas n="nearly 1,000" Mean age: Unclear Follow-up: Unclear Intervention: Abstinence-only vs. Historical control. Data for program participants were compared to data from grade-matched students who attended the same schools one year before the program began. Abstinence By Choice program. Findings: According to Rector 2002, results among female participants favored program participants over controls for "sexual activity rates" (means 5.9% vs. 10.2%). Analyses for male participants also favored the intervention group (means 15.8% vs. 22.8%)
Weed 2005a	Nonrandom assignment of participants. Location: South Carolina n=2,946 at baseline; Analyses for the outcome of interest were limited to 1,469 participants with matched pre-post data who reported being virgins at pretest (1,216 intervention participants, 253 controls). Mean age: Grades 7-9 (ages approx 12-15 yrs) Follow-up: 12 months Intervention: Abstinence-only vs. Usual care. Control schools were selected partially based on their lack of an abstinence education program. Heritage Keepers - 450-minute in-school curriculum based on the Title V, Section 510 (A-H) guidelines. Findings: At 12m follow-up, results significantly favored the intervention group over controls for having initiated sexual activity (means 14.5% vs. 26.5%). "Sexual activity" was assumed to mean vaginal sex
Weed 2005b	Nonrandom assignment of participants. Location: Georgia n=878 at pre-test. After attrition, analyses represent the 464 participants with usable follow-up data on "sexual activity status," who also reported being virgins at pretest. Follow-up: Two cohorts were followed, so follow-up was 12 months or 18 months after pretest. Intervention: Abstinence-only (Choosing the Best) vs. Abstinence-only (Georgia's state-approved abstinence curriculum). Choosing the Best - 2-week in-school curriculum focusing on the risks of teen sex, refusal skills, relationships, self-esteem, alcohol awareness, and parental involvement. Choosing the Best participants received an iteration of the program each year for 2 years. Control participants received 4 sessions from Meeks Heit Totally Awesome Health Course during the first year, and then 4 sessions from Glencoe Health the following year. Findings: "Sexual activity" was assumed to refer to vaginal sex. for the 12-month follow-up sample (n=318), results significantly favored intervention group participants over controls for sexual initiation (means 11.1% vs. 20.9%, $p < 0.001$). For the 18-month follow-up sample (n=146), no significant differences were observed for sexual initiations between intervention participants and controls (means 21.0% vs. 17.4%)
Welmaker 2005	Nonrandom assignment of participants. Location: West Central Georgia n=144 Mean age: 9-14 Follow-up: Immediate post-intervention Intervention: Abstinence-only vs. Usual care. Project Right CHOICES, focuses on character, life skills, and

(*Continued*)

	issues related to health, education, social influences, family, and community. Findings: "Only 2 pregnancies were documented among program participants and only 1 participant indicated at posttest that he had sexual intercourse." Unclear what happened in the control group
Williams 2004	Nonrandom assignment of participants. Report unavailable at the time of this review.
Young 1992	Nonrandom assignment of participants. Location: Arkansas n=209 Age range: Grades 7-8 (ages 12-14) Follow-up: Immediate post-test (within 1 week of program completion). Intervention: Abstinence-only vs. Usual care vs. Nonintervention. Living Smart (middle school component of Sex Can Wait) - 24-lesson in-school curriculum focusing on self-esteem, puberty, decision making, values, communication, dating, goal setting, and life planning. 1 group received the Living Smart curriculum, 1 group received usual care taught by the same facilitators, and 1 group received no instruction. Findings: Students were asked if they had had vaginal intercourse in the past month. the number of students who reported no at pretest and yes at posttest (i.e., who became sexually active) was 0 in the experimental group, 5 in the usual care group, and 3 in the nonintervention control. the number of students who reported yes at pretest and no at posttest (i.e., who stopped being sexually active) was 6 in the Living Smart group, 0 in the usual care group, and 3 in the nonintervention control group. the number of students whose responses were unchanged from pretest to posttest was 77 in the intervention group, 55 in the usual care group, and 60 in the control group

Characteristics of ongoing studies *[ordered by study ID]*

Borawski 2007

Trial name or title	Healthy Teens Building Healthy Schools
Methods	
Participants	1800 youth from 3 ethnically diverse middle schools (ie classes of 2009 and 2010). Participants are enrolled when they start 7th grade (ages 12-13)
Interventions	3 trial arms: 1. Operation Keepsake: Abstinence-only curriculum focusing on abstinence until marriage 2. Draw the Line/Respect the Line: Abstinence-plus curriculum emphasizing abstinence, but also encouraging condom use 3. Attention control group: participants will receive the TEENS general health program and the physical education components of the Planet Health program
Outcomes	Sexual activity, use of protection, engagement in high risk situations
Starting date	2002
Contact information	Case Western Reserve University

Notes	Funding: National Institute of Child Health and Human Development

Jemmott 2006

Trial name or title	Promoting Health Among Teens (PHAT) Project
Methods	
Participants	662 African American grade 6 and 7 students in Pennsylvania, US Mean age=12.0 years 47% Male. 23% reported having had sexual intercourse at baseline, 12% had sexual intercourse in past 3 months 0.3% reported same-gender sexual activity
Interventions	10 trial arms. 1. 8-hour "Making a Difference!" Abstinence-Only Intervention 2. Same 1, with booster sessions. 3. 8-hour Safer-Sex Intervention. 4. Same as 3, with booster sessions. 5. 12-hour Combined Abstinence/Safer Sex. 6. Same as 5, with booster sessions. 7. 8-hour Combined Abstinence/Safer Sex 8. Same as 7, with booster sessions. 9. 8-hour Health Promotion Intervention (Control condition) 10. Same as 9, with booster sessions.
Outcomes	Sexual debut, recent sex, consistent condom use, condom use at last sex
Starting date	Unclear
Contact information	University of Pennsylvania.
Notes	Funding: National Institute of Mental Health Preliminary analyses presented at the International AIDS Conference 2006

Markham 2007

Trial name or title	All About Youth.
Methods	
Participants	1500 students in grade 7, from 38 middle schools in Houston, Texas
Interventions	3 trial arms: 1. Abstinence-only. Based on "It's Your Game ... Keep it real." Includes 22 lessons in 7th and 8th grade focusing on the federal components of abstinence-only education. Includes instruction, group activities, and

Markham 2007 (*Continued*)

	CD-ROMs with individual activities 2. Abstinence-plus. Same format, but abstinence-plus concepts. HIV, STI, and pregnancy prevention program, also based on "It's Your Game ... Keep it real." Includes 22 lessons in 7th and 8th grade focusing on peer relationships, refusal, puberty, body development, consequences of vaginal/oral/anal sex, limits, dating relationships, condoms, and contraceptives. Includes instruction, group activities, and CD-ROMs with individual activities 3. Usual care defined by schools.
Outcomes	Sexual initation, proportion of sexually active students, knowledge, condom use, number of sex partners
Starting date	September 2006
Contact information	University of Texas Houston School of Public Health, Houston TX 77030
Notes	Funding: Centers for Disease Control and Prevention

Miller-Heyl 2007

Trial name or title	Dare to Be You "Care to Wait"
Methods	
Participants	189 youth (aged 12-14) and their families (239 family members), in 2 Colorado sites: Denver and Montezuma County Denver: 92% African-American, 45% of families were comprised of single parents, 69% female parents Montezuma County: 78% white, 22% Hispanic and Native American, 5% of families were comprised of a single parent, and 47% female parents
Interventions	2 trial arms: 1. Dare to Be You "Care to Wait." Based on ecological theory and social learning theory. 22-hour family-based program focusing on skills, self-reponsibility, stress and anger management, family, problem-solving, sex education, and abstinence 2. No treatment.
Outcomes	Percentage of youth participants who are "sexually active," initiation of sex, "secondary virginity"
Starting date	2002
Contact information	Colorado State University.
Notes	Funding: Office of Adolescent Pregnancy Programs. Preliminary findings presented at 2005 and 2007 Abstinence Education Evaluation Conference, US

Nagel 2007

Trial name or title	Positive Choices
Methods	
Participants	510 students in grade 7, from 3 public schools in Toledo, Ohio. Mean age 13.0, 48% male, 47% African-American, 19% white, 23% Hispanic
Interventions	2 trial arms: 1. Positive Choices program. 8 sessions in school, complied with federal abstinence-only definition. Based on Choosing the Best Path and Life curriculum 2. Usual care defined by schools.
Outcomes	Sex in the past 2 months.
Starting date	2005
Contact information	St. Vincent Mercy Medical Center, Toledo OH Ohio State University College of Medicine
Notes	Funding: Office of Population Affairs/Office of Public Health and Service, U.S. Department of Health and Human Services

DATA AND ANALYSES

Comparison 1. Outcomes at each trial's longest follow-up

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 All outcomes	12		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 STI	8		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.2 Pregnancy	9		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.3 Unprotected vaginal sex	5		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.4 Incidence of vaginal sex	7		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.5 No condom use	9		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.6 Sexual initiation	11		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2 Biological outcomes	9		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 STI	8		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2.2 Pregnancy	9		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3 Behavioral outcomes	11		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Unprotected vaginal sex	5		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.2 Incidence of vaginal sex	7		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.3 No condom use	9		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.4 Sexual initiation	11		Odds Ratio (M-H, Random, 95% CI)	Not estimable

Comparison 2. Anderson 1999 (Control: No treatment)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pregnancy, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2 Pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected

Comparison 3. Blake 2001 (Control: Non-enhanced)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of vaginal sex, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2 Sexual initiation, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3 Knowledge of abstinence effectiveness	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected

3.1 Adjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.2 Unadjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
4 Knowledge of pregnancy risk	1	Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 Adjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
4.2 Unadjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable

Comparison 4. Clark 2005 (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of vaginal sex, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2 Incidence of vaginal sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3 Sexual initiation, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4 Sexual initiation, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable

Comparison 5. Goldfarb 1999 (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of unprotected vaginal sex, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2 Incidence of vaginal sex, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3 Did not use condom in past 1 month, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4 Sexual initiation, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable

Comparison 6. Kirby 1997a peer led (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4 Did not use a condom at last sex, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
5 Did not use a condom at last sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
5.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
5.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
6 Sexual initiation, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
6.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
6.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
7 Sexual initiation, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
7.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
7.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable

Comparison 7. Kirby 1997a teacher-led (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4 Did not use a condom at last sex, short-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable

5 Did not use a condom at last sex, long-term	1	Odds Ratio (M-H, Random, 95% CI)	Totals not selected
5.1 Adjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
5.2 Unadjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
6 Sexual initiation, short-term	1	Odds Ratio (M-H, Random, 95% CI)	Totals not selected
6.1 Adjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
6.2 Unadjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
7 Sexual initiation, long-term	1	Odds Ratio (M-H, Random, 95% CI)	Totals not selected
7.1 Adjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable
7.2 Unadjusted for clustering	1	Odds Ratio (M-H, Random, 95% CI)	Not estimable

Comparison 8. Kirby 1997b (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
1.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
2.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3 Did not use a condom at last sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
3.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4 Sexual initiation, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 Adjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable
4.2 Unadjusted for clustering	1		Odds Ratio (M-H, Random, 95% CI)	Not estimable

Comparison 9. Kirby 1997c (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3 Did not use a condom at last sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4 Sexual initiation, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected

Comparison 10. Trenholm 2007a (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3 Incidence of unprotected sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4 Incidence of vaginal sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
5 Lack of condom use at first sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
6 Sexual initation, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected

Comparison 11. Trenholm 2007b (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3 Incidence of unprotected sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4 Incidence of vaginal sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
5 Lack of condom use at first sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
6 Sexual initation, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected

Comparison 12. Trenholm 2007c (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3 Incidence of unprotected sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4 Incidence of vaginal sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected

5 Lack of condom use at first sex, long-term	1	Odds Ratio (M-H, Random, 95% CI)	Totals not selected
6 Sexual initation, long-term	1	Odds Ratio (M-H, Random, 95% CI)	Totals not selected

Comparison 13. Trenholm 2007d (Control: Usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incidence of STI, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
2 Incidence of pregnancy, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
3 Incidence of unprotected sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
4 Incidence of vaginal sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
5 Lack of condom use at first sex, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected
6 Sexual initation, long-term	1		Odds Ratio (M-H, Random, 95% CI)	Totals not selected

ADDITIONAL TABLES

Table 1. Search Strategies for Electronic Databases

Database	Years Searched	Strategy	Total Records
ADOLEC	Inception-2007	Inception - 2005, Total Records Retrieved: 504 Search terms: ABSTINENCE [Words] and SEX [Words] and HIV [Words], SEX [Words] and POSTPONE [Words] or ABSTAIN [Words], VIRGIN [Words] or CHASTITY [Words] or CELIBATE [Words], SEX [Words] and EDUCATION [Words] and HIV [Words]. Examined all records. 2005-2007, Total Records Retrieved: 87 Search terms: sex and education [Palavras] and hiv [Palavras] and 2005 or 2006 or 2007 [País, ano de publicação], abstain and sex, abstain and sex and hiv, postpone and sex, virgin or chaste or chastity or celibate. Examined all records	591

Table 1. Search Strategies for Electronic Databases (*Continued*)

AIDSLINE	1980-2007	1980-2005, Total Unique Records Retrieved: 1450 #1 PT=RANDOMIZED CONTROLLED TRIAL #2 PT=CONTROLLED CLINICAL TRIAL #3 RANDOMIZED CONTROLLED TRIALS #4 RANDOM ALLOCATION #5 DOUBLE BLIND METHOD #6 SINGLE BLIND METHOD #7 PT=CLINICAL TRIAL #8 CLINICAL TRIALS OR CLINICAL TRIALS, PHASE I OR CLINICAL TRIALS, PHASE II OR CLINICAL TRIALS, PHASE III OR CLINICAL TRIALS, PHASE IV OR CONTROLLED CLINICAL TRIALS OR MULTICENTER STUDIES #9 (SINGL* OR DOUBL* OR TREBL* OR TRIPL*) NEAR6 (BLIND* OR MASK*) #10 CLIN* NEAR6 TRIAL* #11 CLIN* NEAR6 TRIAL* #12 PLACEBOS #13 RANDOM* #14 RANDOM* #15 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 #16 ANIMALS NOT (HUMAN AND ANIMALS) #17 #15 NOT #16 #18 SEXUAL ABSTINENCE [MESH] #19 ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED)) #20 #18 OR #19 #21 #17 AND #20 AND PY=1980-2005 2005-2007, Total Unique Records Retrieved: 16 #1 PT=RANDOMIZED	1465

Table 1. Search Strategies for Electronic Databases (*Continued*)

		CONTROLLED TRIAL #2 PT=CONTROLLED CLINICAL TRIAL #3 RANDOMIZED CONTROLLED TRIALS #4 RANDOM ALLOCATION #5 DOUBLE BLIND METHOD #6 SINGLE BLIND METHOD #7 PT=CLINICAL TRIAL #8 CLINICAL TRIALS OR CLINICAL TRIALS, PHASE I OR CLINICAL TRIALS, PHASE II OR CLINICAL TRIALS, PHASE III OR CLINICAL TRIALS, PHASE IV OR CONTROLLED CLINICAL TRIALS OR MULTICENTER STUDIES #9 (SINGL* OR DOUBL* OR TREBL* OR TRIPL*) NEAR6 (BLIND* OR MASK*) #10 CLIN* NEAR6 TRIAL* #11 CLIN* NEAR6 TRIAL* #12 PLACEBOS #13 RANDOM* #14 RANDOM* #15 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 #16 ANIMALS NOT (HUMAN AND ANIMALS) #17 #15 NOT #16 #18 SEXUAL ABSTINENCE [MESH] #19 ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED)) #20 #18 OR #19 #21 #17 AND #20 AND PY=2005-2007	
Bibliomap	1887-2007	1887-2005, Total Unique Records Retrieved: 52 Search terms: EVALUATION AND SEXUAL HEALTH, ABSTINENCE AND SEX AND HIV, SEX AND (POSTPONE OR ABSTAIN), POSTPONE AND SEX, DELAY AND SEX, CHASTITY AND SEX, CHASTITY,	206

Table 1. Search Strategies for Electronic Databases (*Continued*)

		VIRGIN, SEX AND EDUCATION AND HIV. 2005-2007, Total Unique Records Retrieved: 154 abstinen* or abstain* or chastity or chaste or virgin* or celiba* or "sex education"	
BIOSIS	1969-2007	1969-2005, Total Unique Records Retrieved: 1061 (((ABSTINEN* or ABSTAIN* or postpon* or DELAY*) and SEX*) or CHASTITY or CHASTE or VIRGIN* or CELIBA* or (SEX* and EDUCAT*)) and (HIV* or AIDS*) and ((control* or trial* or random* or blind* or (clinical and trial*) or (research and design) or compar* or evaluation) not (animals not human)) 2005-2007, Total Unique Records Retrieved: 185 #1(((ABSTINEN* or ABSTAIN* or postpon* or DELAY*) and SEX*) or CHASTITY or CHASTE or VIRGIN* or CELIBA* or (SEX* and EDUCAT*)) and (PY:BXCD = 2005-2006) #2(HIV* or AIDS*) and (PY:BXCD = 2005-2006) #3((control* or trial* or random* or blind* or (clinical and trial*) or (research and design) or compar* or evaluation) not (animals not human)) and (PY:BXCD = 2005-2006) #4#1 and #2 and #3 and (PY:BXCD = 2005-2006) *Records for 2007 were not yet available in Feb 2007.	1246
Catalog of US Government Publications	1976-2007	1976 - 2005, Total Records Retrieved: 133 Search terms: Abstinence AND sex, (ABSTINEN* OR ABSTAIN* OR DELAY* OR POSTPON*) AND SEX, SEX* AND EDUCATION*, (SEX* AND EDUCATION*) AND (HIV* OR AIDS*). Examined all	428

Table 1. Search Strategies for Electronic Databases (*Continued*)

		records. 2005-2007, Total Records Retrieved: 295 Search terms: Abstinence AND sex, (ABSTINEN* OR ABSTAIN* OR DELAY* OR POSTPON*) AND SEX, SEX* AND EDUCATION*, (SEX* AND EDUCATION*) AND (HIV* OR AIDS*). Examined all records	
CENTRAL	1980-2007	1980-2005, Total Unique Records Retrieved: 74 #1hiv OR hiv-1* OR hiv-2* OR hiv1 OR hiv2 OR (HIV INFECT*) OR (HUMAN IMMUNODEFICIENCY VIRUS) OR (HUMAN IMMUNED-EFICIENCY VIRUS) OR (HUMAN IMMUNE-DEFI- CIENCY VIRUS) OR (HUMAN IM-MUNO-DEFICIENCY VIRUS) OR (HUMAN IMMUN* DEFICIENCY VIRUS) OR (ACQUIRED IMMUN-ODEFICIENCY SYNDROME) OR (ACQUIRED IMMUNEDE-FICIENCY SYNDROME) OR (AC-QUIRED IMMUNO-DEFICIENCY SYNDROME) OR (ACQUIRED IM-MUNE-DE- FICIENCY SYNDROME) OR (AC-QUIRED IMMUN* DEFICIENCY SYNDROME) in All Fields in all prod-ucts #2MeSH descriptor HIV Infections ex-plode all trees in MeSH products #3MeSH descriptor HIV explode all trees in MeSH products #4MeSH descriptor Sexually Transmit-ted Diseases, Viral, this term only in MeSH products #5(#1 OR #2 OR #3 OR #4) #6MeSH descriptor Sexual Abstinence explode all trees in MeSH products #7ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND	212

Table 1. Search Strategies for Electronic Databases (*Continued*)

		(MARRIAGE* OR MARRIED)) in All Fields in all products #8(#6 OR #7) #9(#5 AND #8), from 1980 to 2005 2005-2007, Total Unique Records Retrieved: 138 #1(hiv OR hiv-1* OR hiv-2* OR hiv1 OR hiv2 OR (HIV INFECT*) OR (HUMAN IMMUNODEFICIENCY VIRUS) OR (HUMAN IMMUNED-EFICIENCY VIRUS) OR (HUMAN IMMUNE-DEFI- CIENCY VIRUS) OR (HUMAN IM-MUNO-DEFICIENCY VIRUS) OR (HUMAN IMMUN* DEFICIENCY VIRUS) OR (ACQUIRED IMMUN-ODEFICIENCY SYNDROME) OR (ACQUIRED IMMUNEDE-FICIENCY SYNDROME) OR (AC-QUIRED IMMUNO-DEFICIENCY SYNDROME) OR (ACQUIRED IM-MUNE-DE- FICIENCY SYNDROME) OR (AC-QUIRED IMMUN* DEFICIENCY SYNDROME)), from 2005 to 2007 #2MeSH descriptor HIV Infections explode all trees #3MeSH descriptor HIV explode all trees #4MeSH descriptor Sexually Transmitted Diseases, Viral explode all trees #5MeSH descriptor Sexual Abstinence explode all trees #6ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED)) #7((#1 OR #2 OR #3 OR #4) AND (#5 OR #6)), from 2005 to 2007	
CHID	1985-2005	Search terms: Abstinence, Sex Education, Abstinence and Sex, examined records, most were program manuals or propaganda materials, no visible evaluations, even when using the "professional" audience filter. Examined all records.	0 relevant

Table 1. Search Strategies for Electronic Databases (*Continued*)

		*CHID was taken off-line in September 2006.	
CINAHL	1982-2007	1982-2005, Total Unique Records Retrieved: 4079 (((ABSTINEN* or ABSTAIN* or postpon* or DELAY*) and SEX*) or CHASTITY or CHASTE or VIRGIN* or CELIBA* or (SEX* and EDUCAT*)) and (HIV* or AIDS*) and ((control* or trial* or random* or blind* or (clinical and trial*) or (research and design) or compar* or evaluation) not (animals not human)) 2005-2007, Total Unique Records Retrieved: 143 #1((((ABSTINEN$ or ABSTAIN$ or postpon$ or DELAY$) and SEX$) or CHASTITY or CHASTE or VIRGIN$ or CELIBA$ or (SEX$ and EDUCAT$)) and (HIV$ or AIDS$) and ((control$ or trial$ or random$ or blind$ or (clinical and trial$) or (research and design) or compar$ or evaluation) not (animals not human))).mp. [mp=title, subject heading word, abstract, instrumentation] #2(((ABSTINEN$ or ABSTAIN$ or postpon$ or DELAY$) and SEX$) or CHASTITY or CHASTE or VIRGIN$ or CELIBA$ or (SEX$ and EDUCAT$)).mp. [mp=title, subject heading word, abstract, instrumentation] #3(HIV$ or AIDS$).mp. [mp=title, subject heading word, abstract, instrumentation] #4((control$ or trial$ or random$ or blind$ or (clinical and trial$) or (research and design) or compar$ or evaluat$) not (animals not human)).mp. [mp=title, subject heading word, abstract, instrumentation] #52 and 3 and 4 #6limit 5 to yr="2005 - 2007"	4222
CSA Illumina - includes ASSIA, ERIC, Political Science Abstracts, Social Services Ab-	1963-2007	Sociological Abstracts [1963-], ERIC [1991-], ASSIA [1987-], Political Sci-	1179

Table 1. Search Strategies for Electronic Databases (*Continued*)

stracts, and Sociological Abstracts		ence Abstracts [1975-], and Social Services Abstracts [1979-]. Searched until Febuary 15, 2007 Total Unique Records Retrieved: 1179 1975-2005, Total Unique Records Retrieved: 841 ((ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED))) and (HIV* or AIDS*)) and (randomized controlled trial OR controlled clinical trial OR randomized controlled trials OR random allocation OR double-blind method OR single-blind method OR clinical trial OR clinical trials OR random* OR research design OR comparative study OR evaluation NOT (animals NOT human)) 2005-2007, Total Unique Records Retrieved: 338 (ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED))) and (HIV* or AIDS*) and (randomized controlled trial OR controlled clinical trial OR randomized controlled trials OR random allocation OR double-blind method OR single-blind method OR clinical trial OR clinical trials OR random* OR research design OR comparative study OR evaluation NOT (animals NOT human)), Date Range: 2005-2007	
DARE	1991-2007	1991-2005, Total Unique Records Retrieved: 127 Search terms: HIV, ABSTINENCE, VIRGINITY, VIRGIN, POSTPON* SEX, AIDS 2005-2007, Total Unique Records Retrieved: 122 Search terms: HIV, ABSTINENCE,	249

Table 1. Search Strategies for Electronic Databases (*Continued*)

		VIRGINITY, VIRGIN, POSTPON* SEX, AIDS	
Dissertation Abstracts International (UMI Proquest)	1997-2007	1 KEY(abstinence) or TI(abstinence) or AB(abstinence) 2 #1 and (KEY(hiv) or TI(hiv) or AB (hiv))	64
EMBASE	1974-2007	1980-2005, Total Unique Records Retrieved: 622 #1 ('human immunodeficiency virus infection'/exp) OR ('human immunodeficiency virus'/exp) OR (hiv:ti OR hiv:ab) OR ('hiv-1':ti OR 'hiv-1':ab) OR ('hiv-2':ti OR 'hiv-2':ab) OR ('human immunodeficiency virus':ti OR 'human immunodeficiency virus':ab) OR ('human immuno-deficiency virus':ti OR 'human immuno-deficiency virus':ab) OR ('human immunedeficiency virus':ti OR 'human immunedeficiency virus':ab) OR ('human immune-deficiency virus':ti OR 'human immune-deficiency virus':ab) OR ('acquired immune-deficiency syndrome':ti OR 'acquired immune-deficiency syndrome':ab) OR ('acquired im- munedeficiency syndrome':ti OR 'acquired immunedeficiency syndrome':ab) OR ('acquired immunodeficiency syndrome':ti OR 'acquired immunodeficiency syndrome':ab) OR ('acquired immuno-deficiency syndrome':ti OR 'acquired immuno-deficiency syndrome':ab) #2 (random*:ti OR random*:ab) OR (factorial*:ti OR factorial*:ab) OR (cross?over*:ti OR cross?over:ab OR crossover*:ti OR crossover*:ab) OR (placebo*:ti OR placebo*:ab) OR ((doubl*:ti AND blind*:ti) OR (doubl*:ab AND blind*:ab))) OR (((singl*:ti AND blind*:ti) OR (singl*:ab AND blind*:ab))) OR (assign*:ti OR assign*:ab) OR (volunteer*:ti OR volunteer*:ab) OR ('crossover procedure'/de) OR ('double-blind procedure'/de) OR ('sin-	705

Table 1. Search Strategies for Electronic Databases (*Continued*)

		gle-blind procedure'/de) OR ('randomized controlled trial'/de) OR (allocat*:ti OR allocat*:ab) #3 sexual AND 'abstinence'/exp #4 'abstinence'/de OR abstain* OR chastity OR chaste OR virgin* OR celiba* OR (sex* AND educat*) OR (sex* AND (marriage* OR married)) #5 #3 OR #4 #6 #1 AND #2 AND #5 #7 #1 AND #2 AND #5 AND [1980-2005]/py 2005-2007, Total Unique Records Retrieved: 83 #1exp human immunodeficiency virus infection/ or exp human immunodeficiency virus/ or hiv.ti,ab. or hiv-1.ti,ab. or hiv-2.ti,ab. or human immunodeficiency virus.ti,ab. or human immunodeficiency virus.ti,ab. or human immunedeficiency virus.ti,ab. or human immune-deficiency virus.ti,ab. or acquired immune-deficiency syndrome.ti,ab. or acquired immunedeficiency syndrome.ti,ab. or acquired immunodeficiency syndrome.ti,ab. or acquired immuno-deficiency syndrome.ti,ab. #2(random$ or factorial or cross?over or crossover or placebo).ti,ab. or ((doubl$ and blind$).ti. or (doubl$ and blind$).ab.) or ((singl$ and blind$).ti. or (singl$ and blind$).ab.) or (assign$.ti. or assign$.ab.) or (volunteer$.ti. or volunteer$.ab.) or exp crossover procedure/ or exp double-blind procedure/ or exp single-blind procedure/ or exp randomized controlled trial/ or allocat$.ti,ab. #3exp abstinence/ and sex$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] #4exp abstinence/ or abstain$.mp. or chastity.mp. or chaste.mp. or virgin$.mp. or celiba$.mp. or (sex$ and educat$).mp. or (sex$ and (marriage$ or	

Table 1. Search Strategies for Electronic Databases (*Continued*)

		married)).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] #53 or 4 #61 and 2 and 5 #7limit 6 to yr="2005 - 2007"	
EurasiaHealth Knowledge Multilingual Library	Inception-2007	Inception - 2005, Total Records Retrieved: 452 Search terms: Abstinence, Abstain, Sexual health, AIDS, HIV, HIV Prevention. Examined all records. 2005-2007, Total Records Retrieved: 488 Search terms: Abstinence, Abstain, Sexual health, AIDS, HIV, HIV Prevention. Examined all records	940
HealthPromis	1997-2005	Search Terms: HIV, AIDS, Abstinence. Examined all records.	14
OVID 2005 - includes SIGLE and Global Health	1973-2005	SIGLE (1980 - May 28, 2005) and Global Health (1973 - May 28, 2005) Total Unique Records Retrieved: 683 (((ABSTINEN* or ABSTAIN* or postpon* or DELAY*) and SEX*) or CHASTITY or CHASTE or VIRGIN* or CELIBA* or (SEX* and EDUCAT*)) and (HIV* or AIDS*) and ((control* or trial* or random* or blind* or (clinical and trial*) or (research and design) or compar* or evaluation) not (animals not human)) *Please see Ovid 2007 search for Global Health 2005-2007. SIGLE was taken off line in 2005	683
OVID 2007 - includes AMED, BNI, HMIC, and Global Health	2005-2007	AMED [2005-], BNI [2005-], HMIC [2005-], and Global Health [2005-], Searched February 15, 2007 Total Unique Records Retrieved: 118 #1(((ABSTINEN$ or ABSTAIN$ or postpon$ or DELAY$) and SEX$) or CHASTITY or CHASTE or VIR-	118

Table 1. Search Strategies for Electronic Databases (*Continued*)

		GIN$ or CELIBA$ or (SEX$ and EDUCAT$)).mp. [mp=ab, hw, ti, ot, bt] #2(HIV$ or AIDS$).mp. [mp=ab, hw, ti, ot, bt] #3((control$ or trial$ or random$ or blind$ or (clinical and trial$) or (research and design) or compar$ or evaluat$) not (animals not human)).mp. [mp=ab, hw, ti, ot, bt] #41 and 2 and 3 #5limit 4 to yr="2005 - 2007"	
PAIS 2007	2005-2007	1972-2005, see WebSpirs 2005 - February 15, 2007 Total Unique Records Retrieved: 9 (ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED))) and (hiv* or aids*) and (randomized controlled trial OR controlled clinical trial OR randomized controlled trials OR random allocation OR double-blind method OR single-blind method OR clinical trial OR clinical trials OR random* OR research design OR comparative study OR evaluation NOT (animals NOT human))	9
PsycINFO	1887-2007	1887-2005, Total Unique Records Retrieved: 1146 ((ABSTINEN* or ABSTAIN* or postpon* or DELAY*) and SEX*) or CHASTITY or CHASTE or VIRGIN* or CELIBA* or (SEX* and EDUCAT*)) and (HIV* or AIDS*) and ((control* or trial* or random* or blind* or (clinical and trial*) or (research and design) or compar* or evaluation) not (animals not human)) 2005-2007, Total Unique Records Retrieved: 99 #1((((ABSTINEN$ or ABSTAIN$ or postpon$ or DELAY$) and SEX$) or CHASTITY or CHASTE or VIRGIN$ or CELIBA$ or (SEX$ and ED-	1245

Table 1. Search Strategies for Electronic Databases (*Continued*)

		UCAT$)) and (HIV$ or AIDS$) and ((control$ or trial$ or random$ or blind$ or (clinical and trial$) or (research and design) or compar$ or evaluation) not (animals not human))).mp. [mp=title, abstract, heading word, table of contents, key concepts] #2limit 1 to yr="2005 - 2007"	
PubMed	1980-2007	1980-2005, Total Unique Records Retrieved: 5164 #1 Search HIV Infections[MeSH] OR HIV[MeSH] OR hiv[tw] OR hiv-1*[tw] OR hiv-2*[tw] OR hiv1[tw] OR hiv2[tw] OR hiv infect*[tw] OR human immunodeficiency virus[tw] OR human immunedeficiency virus[tw] OR human immuno-deficiency virus[tw] OR human immune-deficiency virus[tw] OR ((human immun*) AND (deficiency virus[tw])) OR acquired immunodeficiency syndrome[tw] OR acquired immunedeficiency syndrome[tw] OR acquired immuno-deficiency syndrome[tw] OR acquired immune-deficiency syndrome[tw] OR ((acquired immun*) AND (deficiency syndrome[tw])) OR "Sexually Transmitted Diseases, Viral"[MeSH:NoExp] #2 Search SEXUAL ABSTINENCE [MESH] #3 Search ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED)) #4 Search #2 OR #3 #5 Search randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh] OR clinical trial [pt] OR clinical trials [mh] OR ("clinical trial" [tw]) OR ((singl* [tw] OR doubl* [tw] OR trebl* [tw] OR tripl* [tw]) AND	5759

Table 1. Search Strategies for Electronic Databases (*Continued*)

		(mask* [tw] OR blind* [tw])) OR (placebos [mh] OR placebo* [tw] OR random* [tw] OR research design [mh: noexp] OR comparative study [mh] OR evaluation studies [mh] OR follow-up studies [mh] OR prospective studies [mh] OR control* [tw] OR prospectiv* [tw] OR volunteer* [tw]) NOT (animals [mh] NOT human [mh]) #6Search #1 AND #4 AND #5 Field: All Fields, Limits: Publication Date from 1980 to 2005 2005-2007, Total Unique Records Retrieved: 595 #1 Search HIV Infections[MeSH] OR HIV[MeSH] OR hiv[tw] OR hiv-1*[tw] OR hiv-2*[tw] OR hiv1[tw] OR hiv2[tw] OR hiv infect*[tw] OR human immunodeficiency virus[tw] OR human immunedeficiency virus[tw] OR human immuno-deficiency virus[tw] OR human immune-deficiency virus[tw] OR ((human immun*) AND (deficiency virus[tw])) OR acquired immunodeficiency syndrome[tw] OR acquired immunedeficiency syndrome[tw] OR acquired immunodeficiency syndrome[tw] OR acquired immune-deficiency syndrome[tw] OR ((acquired immun*) AND (deficiency syndrome[tw])) OR "Sexually Transmitted Diseases, Viral"[MeSH:NoExp] Limits: Publication Date from 2005 to 2007 #2 Search SEXUAL ABSTINENCE [MESH] Limits: Publication Date from 2005 to 2007 #3 Search ABSTINENCE OR ABSTAIN* OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*) OR (SEX* AND (MARRIAGE* OR MARRIED)) Limits: Publication Date from 2005 to 2007 #4 Search #2 OR #3 Limits: Publica-	

Table 1. Search Strategies for Electronic Databases (*Continued*)

		tion Date from 2005 to 2007 #5 Search randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh] OR clinical trial [pt] OR clinical trials [mh] OR ("clinical trial" [tw]) OR ((singl* [tw] OR doubl* [tw] OR trebl* [tw] OR tripl* [tw]) AND (mask* [tw] OR blind* [tw])) OR (placebos [mh] OR placebo* [tw] OR random* [tw] OR research design [mh: noexp] OR comparative study [mh] OR evaluation studies [mh] OR follow-up studies [mh] OR prospective studies [mh] OR control* [tw] OR prospectiv* [tw] OR volunteer* [tw]) NOT (animals [mh] NOT human [mh]) Limits: Publication Date from 2005 to 2007 #6 Search #1 AND #4 AND #5 Limits: Publication Date from 2005 to 2007	
SciSearch (Web of Knowledge) : Biomedical abstracts	1974-2007	1974-2005, Total Unique Records Retrieved: 21 #1 topic=(((ABSTINEN* OR ABSTAIN*) and SEX*) OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*)) Database=Web of Science; Timespan=Latest 5 Years #2 topic=(HIV* OR AIDS*) Database=Web of Science; Timespan=Latest 5 Years #3 topic=((randomized SAME controlled) OR control* OR trial* OR random* OR double-blind* OR single-blind* OR (clinical AND trial*) OR (research SAME design) OR comparative OR evaluation NOT (animals NOT human)) Database=Web of Science; Timespan=Latest 5 Years #4 topic=((((ABSTINEN* OR ABSTAIN*) and SEX*) OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*)) and (HIV* OR AIDS*) and	87

Table 1. Search Strategies for Electronic Databases (*Continued*)

		((randomized SAME controlled) OR control* OR trial* OR random* OR double-blind* OR single-blind* OR (clinical AND trial*) OR (research SAME design) OR comparative OR evaluation NOT (animals NOT human))) 2005-2007, Total Unique Records Retrieved: 66 #1 topic=(((ABSTINEN* OR ABSTAIN*) and SEX*) OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*)) Database=Web of Science; Timespan=Latest 5 Years #2 topic=(HIV* OR AIDS*) Database=Web of Science; Timespan=Latest 5 Years #3 topic=((randomized SAME controlled) OR control* OR trial* OR random* OR double-blind* OR single-blind* OR (clinical AND trial*) OR (research SAME design) OR comparative OR evaluation NOT (animals NOT human)) Database=Web of Science; Timespan=Latest 5 Years #4 topic=((((ABSTINEN* OR ABSTAIN*) and SEX*) OR CHASTITY OR CHASTE OR VIRGIN* OR CELIBA* OR (SEX* AND EDUCAT*)) and (HIV* OR AIDS*) and ((randomized SAME controlled) OR control* OR trial* OR random* OR double-blind* OR single-blind* OR (clinical AND trial*) OR (research SAME design) OR comparative OR evaluation NOT (animals NOT human))) Database=Web of Science; Timespan=Latest 5 Years Examined records from 2005-2007.	
TROPHI	Inception-2007	Inception - 2005, Total Records Retrieved: 176 Search terms: abstinence AND sex AND HIV, sex and (postpone or abstain), postpone AND sex, delay AND	376

Table 1. Search Strategies for Electronic Databases (*Continued*)

		sex, chastity, virgin, sex education AND HIV, HIV OR AIDS. Examined all records. 2005-2007, Total Records Retrieved: 200 Search terms: abstinence and sex, virgin, postpone sex, abstain, HIV. Examined all records from 2005-2007	
WEBSPIRS 2005 - includes AMED, BNI, RCN Journals, HMIC, PAIS, and SERFILE	1972-2005	AMED [1985-], BNI [1985-], RCN [1985-1996], HMIC [1983-], SERFILE [Inception-], and PAIS [1972-], Searched May 28, 2005 Total Unique Records Retrieved: 94 (((ABSTINEN* or ABSTAIN* or postpon* or DELAY*) and SEX*) or CHASTITY or CHASTE or VIRGIN* or CELIBA* or (SEX* and EDUCAT*)) and (HIV* or AIDS*) and ((control* or trial* or random* or blind* or (clinical and trial*) or (research and design) or compar* or evaluation) not (animals not human)) *For 2005-2007 searches of AMED, BNI, and HMIC, see the search of OVID. For 2005-2007 searches of PAIS, see PAIS. RCN was searchable via BNI in 2007	94

WHAT'S NEW

Last assessed as up-to-date: 21 August 2007.

Date	Event	Description
10 November 2008	Amended	Converted to new review format.

HISTORY

Protocol first published: Issue 3, 2005

Review first published: Issue 4, 2007

CONTRIBUTIONS OF AUTHORS

All authors contributed equally to conceiving and designing the review, screening studies for inclusion, appraising the quality of papers, extracting data from papers, drafting and editing the protocol and full review, and interpreting the data.

KU coordinated the review, undertook searches of electronic databases that were not searched by the Cochrane HIV/AIDS Group Trial Search Coordinator (ie all but CENTRAL, PubMed, AIDSLINE, and EMBASE in 2005, all databases in 2007), carried out the email search for unpublished literature, and organized the retrieval of papers.

KU and DO wrote to authors of papers for additional information.

KU and PM extracted and entered data on trial results into RevMan.

All authors have approved this manuscript.

At the time of this work, KU was a Research Officer at the Centre for Evidence-Based Intervention, University of Oxford.

DECLARATIONS OF INTEREST

None

SOURCES OF SUPPORT

Internal sources

- Centre for Evidence-Based Intervention, University of Oxford, UK.
- Department of Social Policy and Social Work, University of Oxford, UK.

External sources

- No sources of support supplied

INDEX TERMS

Medical Subject Headings (MeSH)

*Developed Countries; *Program Evaluation; *Sexual Abstinence; Disease Outbreaks [*prevention & control]; HIV Infections [*prevention & control]; Randomized Controlled Trials as Topic; Safe Sex; United States

MeSH check words

Humans