MAY 2014

ASPE RESEARCH BRIEF

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION

OFFICE OF HUMAN SERVICES POLICY - U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

USING EFFECT SIZES TO INFORM POLICY AND PRACTICE

Evidence from the HHS Teen Pregnancy Prevention Evidence Review

In fall 2009, the U.S. Department of Health and Human Services (HHS) launched a systematic review of the research literature on programs to prevent teen pregnancy, sexually transmitted infections (STIs), and associated sexual risk behaviors. Findings have been used in part to identify programs with evidence of effectiveness in reducing these outcomes. To help inform researchers, policymakers, and practitioners about the size of the effects produced by these programs, this research brief summarizes an ongoing effort to collect and report program effect size information from the reviewed studies. Findings indicate substantial variation in effect sizes across programs, but also a clear need for improved standards and reporting of effect size information in teen pregnancy prevention research.

Research in the teen pregnancy prevention literature has identified a broad range of programs with evidence of effectiveness in reducing teen pregnancy, STIs, and associated sexual risk behaviors. An ongoing systematic review of the literature conducted for HHS by Mathematica Policy Research and its partner, Child Trends, has identified 31 different programs with demonstrated evidence of effectiveness, based on a detailed assessment of research released from 1989 through early 2011. These programs range from short one-on-one clinical or counseling interventions to broad multi-year youth development programs. Many of the supporting research studies are based on rigorous randomized controlled trials that provide a sound basis for estimating program effects.

However, relatively less is known about the size or magnitude of the effects produced by these programs. The supporting research studies demonstrate that youth who are offered the programs have more favorable outcomes than

ABOUT THIS RESEARCH BRIEF

This ASPE Research Brief, which describes the effect sizes of evidence-based teen pregnancy prevention programs, was written by Mary Terzian and Vanessa Sacks of Child Trends. The analysis was conducted as part of an ongoing systematic review of the teen pregnancy prevention literature conducted by Mathematica Policy Research, in partnership with Child Trends, under contract with HHS.

Office of the Assistant Secretary for Planning and Evaluation

Office of Human Services Policy

US Department of Health and Human Services

Washington, DC 20201

youth who are not. However, studies do not always report the size of these differences or how the measured effects compare across programs.

To help fill this gap, this research brief summarizes findings from an ongoing effort to systematically collect and report program effect size information for the 31 programs featured in the HHS Teen Pregnancy Prevention Evidence Review. All of the programs have been determined to meet the review criteria for demonstrated evidence of effectiveness in reducing teen pregnancy, STIs, or associated sexual risk behaviors. However, this brief provides the first information on effect sizes for these programs.

What is an Effect Size?

When discussed in the context of a program evaluation, an effect size represents the magnitude of a program effect for a particular outcome. In other words, it is an estimate of the size of the effect. Effect sizes are useful for many purposes. For program providers, effect sizes can be used to select a program, monitor program performance, and assess whether a program is having its intended effect. Effect sizes can also provide useful information to program sponsors and funders—for example, when assessing a program's likelihood to return on an investment or when choosing from among many different potential programs to support. Effect sizes also play an important technical role in evaluation research—for example, when calculating statistical power in the planning stages of an evaluation, or when comparing or combining program impact estimates across studies as in a meta-analysis (Hill, Bloom, Black, & Lipsey, 2008).

Effect sizes provide unique information not captured by other statistics commonly reported in published reports or journal articles. In particular, although researchers typically report whether a difference in outcomes between the treatment and comparison groups is "statistically significant," this information alone does not indicate the practical size of the reported difference. Effect size estimates complement statistical significance reporting by expressing the size or magnitude of the difference between the treatment and comparison groups. Compared with statistical significance tests, effect sizes are relatively less sensitive to differences in study sample size.

How to calculate an effect size

The reported magnitude of an effect may be based on the difference in means or proportions between treatment and comparison groups (or, for randomized studies, between treatment and control groups). This difference could be based on a point-in-time estimate (comparing posttest means, for example); a difference-in-difference estimate (comparing the relative change in the treatment group to the relative change in the comparison group from pretest to posttest); or it could be based on an estimate of the rate of change over time (such as from growth curve analyses).

Effect sizes can be presented in standardized units or in the original units of the outcome measure (unstandardized units). For continuous outcomes, common standardized measures include standardized mean differences, correlations, or variance explained statistics. For dichotomous or binary measures (e.g., yes/no questions), common standardized measures include ratios of odds (odds ratios, or "OR") and ratios of proportions (risk ratios, or "RR"). Common

measures of unstandardized effect sizes include percentage point differences (e.g., the difference in the percent of youth in the treatment group reporting recent sex and the percent of the comparison group reporting recent sex) and raw mean differences (e.g., the difference in the average number of lifetime sexual partners between the treatment group and the comparison group).

Data and Methods

This brief is based on information for the 31 programs identified in the HHS Teen Pregnancy Prevention Evidence Review as having demonstrated evidence of effectiveness in reducing teen pregnancy, STIs, or associated sexual risk behaviors. For these programs, we reviewed the supporting impact studies (n = 35) in an effort to collect effect size estimates for all program impact estimates in the following outcome domains: sexual activity, condom use, pregnancy, and STIs. For continuous outcomes, we sought estimates of standardized mean differences expressed as Cohen's *d*. For dichotomous outcomes, we sought estimates of odds ratios.

Across the 35 studies included in the analysis, we identified a total of 339 separate impact estimates for the outcomes of interest. Over a third of these impact estimates were reported as statistically significant by the study authors. The majority of these estimates (200) were for dichotomous outcomes. A smaller number (139) were for continuous outcomes. The larger number of dichotomous outcomes reflects a common focus in this literature on using questions warranting a "yes/no" (dichotomous) response (such as "Have you ever been pregnant?" or "Have you ever had sex?").We sought information for all relevant impact estimates regardless of the level of statistical significance reported in the study.

Many studies did not directly report the relevant effect size information of interest (Cohen's *d* for continuous outcomes and odds ratios for dichotomous outcomes). In some cases, we addressed this limitation by calculating an effect size from other information provided in the report or article. In other cases, we were able to obtain or

INTERPRETING EFFECT SIZE ESTIMATES

There are no universally accepted guidelines for interpreting the clinical or practical significance of effect size estimates. Simply describing the magnitude of an effect size may not give the full picture of its practical or clinical value, since a seemingly "small" effect size on an important outcome may have large and meaningful practical effects. Simple descriptions of effect size estimates may also be open to subjective interpretation. Cohen (1988) recommends that a given effect size should be interpreted in relation to other effect sizes obtained within the same field of practice. However, in the field of teen pregnancy prevention research, few such benchmark estimates currently exist.

Given these uncertainties, researchers often rely on standard conventions or "rules of thumb" to interpret effect size estimates. In particular, Cohen (1988) developed guidelines that many researchers still use today to interpret effect size estimates. For continuous outcomes, this convention defines a standardized mean difference of 0.2 as a "small" effect size; 0.5 as a "medium" effect size; 0.8 as a "large" effect size; and 1.0 as a "very large" effect size.

For dichotomous outcomes, the convention defines the following as cutoff points for odds ratios: 1.5 (small effect size); 2.5 (medium effect size); and 4.3 (large effect size). Such guidelines must be interpreted with caution, however, given differences in the clinical and practical significance of outcomes in different fields.

calculate an effect size based on additional information provided through follow-up with the study author. There were three studies (out of 35) for which we could not calculate or obtain any effect sizes for the outcomes of interest. However, for many other studies were not always able to produce an effect size for every relevant outcome in a particular study. Ultimately, we were successful in either obtaining or calculating effect sizes for just over half (180) of the 339 impact estimates considered for the analysis. We weigh the implications of this challenge in the discussion section.

We found that effect size estimates were easiest to obtain or calculate for dichotomous outcomes. Of the 200 impact estimates for dichotomous outcomes, sufficient data were available to obtain or calculate an effect size (odds ratio) for 81% (162 estimates). By contrast, of the impact estimates for 139 continuous outcomes, sufficient data were available to obtain or calculate an effect size (Cohen's *d*) for only 12% (17 estimates). This was a surprising finding given the common use of Cohen's *d* as an effect size measure in other areas of prevention research (Mindel & Hoefer, 2006). It was harder to obtain all the information needed to calculate a Cohen's *d* largely because more information is needed, while an odds ratio can be calculated using two proportions. For a more detailed description of the methods used for this brief, see Appendix A.

Results: What Did We Discover About Effect Sizes in This Review?

We present results separately for each of four outcome domains: sexual activity, condom use, pregnancy, and STIs. Within each domain, we summarize the measures examined, data availability, and range of effect size estimates obtained or calculated. We do not report mean effect sizes for each domain, given that we did not conduct a meta-analysis to control for variation in program models and evaluation designs. An example source of variation is that some program models were designed to be implemented over multiple years, while others were delivered across a single school year. In this example, even if both studies measured condom use, we concluded that the effect sizes calculated from the findings would represent the results of very different treatment exposures. To view the full list of effect sizes obtained or calculated as part of this analysis go to http://tppevidencereview.aspe.hhs.gov.

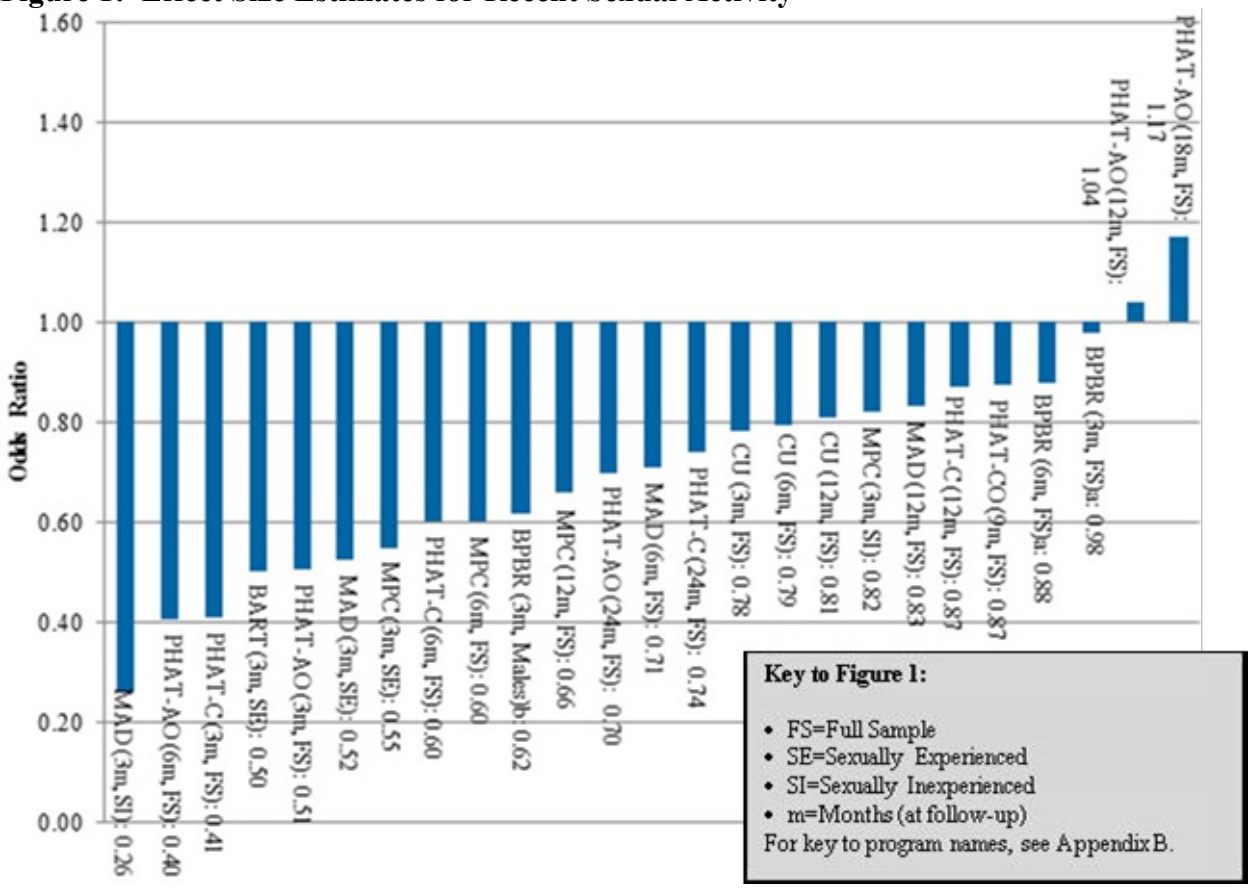
Sexual Activity

The majority of studies included in the analysis examined program impacts on at least one measure of sexual activity. The most common measures examined were recent sexual activity (11 studies), number of sexual partners (10 studies), and sexual initiation (10 studies). Other outcomes examined include frequency of sexual activity (8 studies), having had multiple sexual partners (7 studies), and ever having had sex (2 studies). However, not every study reported enough information to calculate an effect size for all impact estimates.

For sexual initiation, the estimated odds ratios range widely, from a low of 0.17 to a high of 2.35. An odds ratio greater than 1 indicates that the treatment group had higher odds than the comparison group of initiating sex, while an odds ratio lower than one indicates the opposite. See the call out box "Interpreting Effect Size Estimates" for more detail on the guidelines for interpreting the size of an odds ratio. The range of effect sizes for sexual initiation is based on 22 impact estimates from 7 studies with available data. Odds ratios were not available for two additional studies that included sexual initiation as an outcome measure. Of the 19 available

estimates, the large majority (15) of odds ratios were below 1.0, indicating lower rates of sexual initiation for the treatment group. However, effect sizes for 4 of the 19 available estimates were greater than or equal to 1.0, indicating no favorable program effect. The lowest odds ratio of 0.17 indicates that the treatment group odds of initiating sexual activity were less than one-fifth the odds for the comparison group. Of the 19 available estimates, some were calculated separately for males and females whereas others were calculated for males and females together.

Figure 1: Effect Size Estimates for Recent Sexual Activity*



* Defined as vaginal intercourse in the past three months. A list of the study citations corresponding to each program is listed in Appendix B.

a. Jemmott, Jemmott, Fong, & McCaffree (1999).

b. Jemmott, Jemmott, &, Fong (1992).

For recent sexual activity, the variation in effect sizes is somewhat narrower, with odds ratios ranging from 0.26 to 1.17 (Figure 1). This range is based on 25 impact estimates from 8 studies with available data that define recent sexual activity as having had vaginal intercourse in the past three months. Some of the estimates are based on the full study sample whereas others are subgroup estimates. The estimates also vary with respect to length of follow up. For example, for a study of the abstinence-focused *Making a Difference* program, odds ratios range from 0.26 for a subgroup estimate measured at a 3-month follow-up survey to 0.86 for a full sample estimate measured at a 12-month follow-up. Such differences in the analysis sample and length of follow-up may contribute to the observed differences in odds ratios. Three other studies measured recent sexual activity over the past 19 weeks to one year, with odds ratios ranging from 0.49 to 1.10.

Calculating effect sizes for other measures of sexual activity presented several challenges. For one, we were often unable to obtain or calculate effect sizes for continuous measures such as the number of sexual partners or frequency of sex. For example, although 10 studies examined program impacts on a measure of the number of sexual partners, we were able to obtain or calculate effect size estimates for only two of these studies. We also found a lack of consistency in measurement across studies. For example, whereas some studies measured the number of sexual partners as a continuous or count variable, others defined the outcome as a dichotomous or categorical variable (such as categories for zero partners, one partner, and multiple partners). This lack of consistency limited our ability to make effect size comparisons even when a number of estimates were available.

Condom Use

Condom use is a particularly challenging measure for calculating and comparing effect sizes. Just over half of the studies included in the review examined program impacts on at least one measure of condom use, making it one of the most commonly-measured outcomes. However, the studies do not always measure the construct in similar ways. Some studies focused on continuous measures of condom use, such as the frequency of sex without a condom, whereas other studies used dichotomous measures, such as condom use at last sex or consistent condom use. For the purpose of this analysis, we present the dichotomous measures because effect size estimates are not available for many of the continuous measures.

Table 1: Condom Use at Last Sex

Program	Citation	Follow-Up Time	Odds Ratio*
Be Proud! Be Responsible!	Jemmott, Jemmott, Fong, & Morales (2010)	3 months	1.22 (SE)
		6 months	1.12 (SE)
		12 months	1.05 (SE)
Horizons	DiClemente et al. (2009)	6 months	1.40
		12 months	1.53
Safer Choices	Kirby, Baumler, & Coyle (2011)	12 months	1.63 (SE)
		24 months	1.44 (SE)
		31 months	0.88 (SE)
Safer Sex	Shrier et al. (2001)	1 month	0.85
SiHLE	DiClemente et al. (2004)	6 months	5.08 (females)
		12 months	3.32 (females)

* OR > 1.0 indicates the odds of condom use at last sex among the treatment group are higher than the odds of condom use at last sex among the comparison group. Estimates are for a co-ed, full sample except where noted: SE = subgroup of sexually experienced at baseline.

For studies examining program impacts on condom use at last sex, the estimated odds ratios range widely, from 0.85 to 5.08 (Table 1). This range is based on 11 impact estimates from five studies with available data. The majority of odds ratios (82%) are greater than 1.0, indicating higher odds of condom use at last sex among the treatment group. However, two of the estimates indicate higher odds of condom use for the comparison group (OR < 1.0). Among the four

studies measuring this outcome at multiple follow-up periods, three show a pattern of diminishing effect sizes over time, evidence of a weakening program effect.

Four studies examined program impacts on a dichotomous measure of condom use consistency—that is, a measure of whether participants used condoms not just the last time but *every* recent time they had sex. Some studies examined consistency over a relatively short period of a month or less, whereas others focused on longer periods of three to six months. Odds ratios for this outcome range from 1.05 to 2.48 based on a sample of 14 impact estimates (Table 2). All of the odds ratios are above 1.0, indicating higher odds of consistent condom use among youth in the treatment group. Unlike results for the measure of condom use at last sex, we find no consistent pattern of diminishing effect sizes over time for consistent condom use. The estimated effect sizes become smaller over time for some studies but increase for others.

Table 2: Consistent Condom Use

Program	Citation	Outcome Reference Period	Follow-Up Time	Odds Ratio*
Be Proud! Be Responsible!	Jemmott, Jemmott, Fong, & Morales (2010)	Past 3 months	3 months	1.25 (SE)
			6 months	1.12 (SE)
			12 months	1.28 (SE)
¡Cuídate!	Villarruel, Jemmott, & Jemmott (2006)	Past 3 months	3 months	2.15 (SE)
			6 months	2.01 (SE)
			12 months	1.92 (SE)
Horizons	DiClemente et al. (2009)	Past 14 days	6 months	1.05
			12 months	1.55
		Past 60 days	6 months	1.19
			12 months	1.58
SiHLE	DiClemente et al. (2004)	Past 30 days	6 months	1.77 (females)
			12 months	2.23 (females)
		Past 6 months	6 months	2.48 (females)
			12 months	2.14 (females)

* OR > 1.0 indicates the odds of consistent condom use among the treatment group are higher than the odds of consistent condom use among the comparison group. Estimates are for a co-ed, full sample except where noted: SE = subgroup of sexually experienced at baseline.

Other studies examined program impacts on condom use by asking respondents questions about unprotected sex. In particular, for three of the studies included in our analysis, we obtained or calculated odds ratios for a measure indicating whether participants had sex without a condom in the past three months. The estimated odds ratios for this outcome range odds ratio below 1.0 indicates program impacts favorable to the treatment group, and all but three of the reported impact estimates meet this cutoff. More than half the estimates (71%) are for subgroups based on youth's sexual experience at baseline, whereas the others are based on the full study sample.

Table 3: Sex without a Condom

Program	Citation	Follow-Up Time	Odds Ratio*
Making A Difference	Jemmott, Jemmott, & Fong (1998)	3 months	0.63 (SE)
			0.03 (SI)
		6 months	0.80 (SE)
			0.66 (SI)
		12 months	1.02 (SE)
			1.18 (SI)
Making Proud Choices	Jemmott, Jemmott, & Fong (1998)	3 months	0.20 (SE)
			0.46 (SI)
		6 months	0.48 (SE)
			0.83 (SI)
		12 months	0.23 (SE)
			0.66 (SI)
Promoting Health Among Teens-Abstinence-Only	Jemmott, Jemmott, & Fong (2010)	3 months	0.24
		6 months	0.08
		12 months	0.95
		18 months	1.14
		24 months	0.97

* OR < 1.0 indicates that the treatment group had lower odds of having sex without a condom than the comparison group. Estimates are for a co-ed, full sample except where noted: SE = subgroup of sexually experienced at baseline; SI = subgroup of sexually inexperienced at baseline.

Pregnancy

Relatively few studies examined long-term program impacts on pregnancy (Table 4). Data were available to calculate program effect sizes on pregnancy from six studies. A seventh study also measured program impacts on pregnancy but was excluded from our analysis because of missing effect size information.

Despite this relatively small sample, one advantage of focusing on pregnancy as an outcome is greater consistency in measurement. Most studies measure pregnancy with simple dichotomous measures of ever been pregnant (for females) or gotten someone pregnant (for males). The relative simplicity and consistency of these measures makes pregnancy an easier outcome for which to analyze and compare effect sizes than sexual activity or condom use, which are not always measured in consistent ways.

Among the six studies with available data, the estimated odds ratios for pregnancy range from 0.31 to 1.17 (Table 4) based on a sample of 10 impact estimates. All but one of the odds ratios were below 1.0, indicating lower pregnancy rates for the treatment than comparison group. The smallest estimated odds ratio of 0.31 indicates that the odds of having a pregnancy or causing a pregnancy among the treatment group participants were less than a third of the odds for comparison group participants.

Table 4: Pregnancy

Program	Citation	Follow-Up Time	Odds Ratio*
CAS-Carrera	Philliber et al. (2002)	During program (3 years from baseline)	0.31 (females)
		During program (3 years from baseline)	1.17 (males)
Raising Healthy Children	Hawkins et al. (1999)	Age 18 (10 years from baseline)	0.58
	Lonczak et al. (2002)	Age 21 (13 years from baseline)	0.50 (females)
		Age 21 (13 years from baseline)	0.95 (males)
	Hawkins et al. (2008)	Age 24 (16 years from baseline)	0.68
		Age 27 (19 years from baseline)	0.69
SiHLE	DiClemente et al. (2004)	6 months (after posttest)	0.38 (females)
		12 months (after posttest)	0.74 (females)
Teen Outreach Program	Sikkema et al. (2005)	0 months (posttest)	0.41

* OR < 1.0 indicates that the treatment group had lower odds of experiencing or caused a pregnancy than the comparison group. Estimates are for a co-ed, full sample except where noted.

Half of the available estimates are from a long-term study of *Raising Healthy Children*, a 5-year, school-wide intervention for elementary school students. The study conducted several long-term follow ups with study participants in their late teens and 20s, allowing for a long-term assessment of program impacts on pregnancy rates. The study findings show a pattern of diminishing program impacts with age: an odds ratio of 0.58 when the participants were age 18, 0.68 at age 24, and 0.69 at age 27. This pattern suggests the program had an effect in changing the timing of pregnancy, but not necessarily in changing the rates of ever becoming pregnant.

Sexually Transmitted Infections

Six studies assessed program impacts on measures of sexually transmitted infections (STIs). Unlike measures of pregnancy, STI measures are not always defined consistently across studies. Some studies focus on the prevalence of specific STIs such as Chlamydia, Gonorrhea, or Trichomoniasis, whereas other studies focus on more general measures of infection with any STI. The estimated odds ratios range from 0.14 to 1.06 (Table 5) based on a sample of 13 available impact estimates. The majority of the estimates were below 1.0, indicating lower rates of STI infection among treatment group members. Two studies examined program impacts separately for three specific STIs (Chlamydia, Gonorrhea, or Trichomoniasis), in some cases finding different results. Such findings caution against generalizing effect size estimates across measures of STIs.

Table 5: Sexually Transmitted Infections

Program	Citation	Follow-Up Time	STI Measured	Odds Ratio*
Horizons	DiClemente et al. (2009)	12 months	Chlamydia Gonorrhea Trichomoniasis	0.68 1.01 1.00
Raising Healthy Children	Lonczak et al. (2002)	Age 21 (13 years from baseline)	Any STI	0.67
	Hawkins et al. (2008)	Age 24 (16 years from baseline)	Any STI	0.52
		Age 27 (19 years from baseline)	Any STI	0.55
SiHLE	DiClemente et al. (2004)	12 months	Chlamydia Gonorrhea Trichomoniasis	0.17 (females) 0.14 (females) 0.37 (females)
Sisters Saving Sisters	Jemmott, Jemmott, Braverman, & Fong (2005)	6 months	Any STI	1.06 (females)
		12 months	Any STI	0.5 (females)
What Could You Do?	Downs et al. (2004)	6 months	Chlamydia	0.36
		6 months	Any STI	0.36

* OR < 1.0 indicates that the treatment group had lower odds of a sexually transmitted infection than the comparison group. Estimates are for a co-ed, full sample except where noted.

Discussion

The findings presented in this brief show wide variation in effect sizes across teen pregnancy prevention programs. Within each of the four outcome domains examined (sexual activity, condom use, pregnancy, and STIs), the estimated effect sizes vary widely, from strongly favorable effects for the treatment group to null or non-significant effects favoring the comparison group. For outcomes on which odds ratios below 1.0 favor the treatment group, we found odds ratios ranging from 0.17 to 2.35 for sexual initiation, 0.13 to 1.17 for recent sex, 0.31 to 1.17 for pregnancy, and 0.14 to 1.06 for STIs. For outcomes on which odds ratios larger than 1.0 favor the treatment group, we found odds ratios ranging from 1.05 to 2.48 for consistent condom use and 0.85 to 5.08 for condom use at last sex.

This large variation in effect sizes may partly reflect the relatively small number of estimates available and limited basis for making comparisons across programs. Many of the studies included in our analysis do not report standardized effect sizes or the information needed to calculate such effects. In addition, when contacted, study authors were often unable to provide additional information on program effect sizes, particularly for older studies for which data have been lost or archived. As a result, we were successful in obtaining or calculating effect sizes for just over half of the 339 impact estimates reviewed—including only 12% of the impact estimates for continuous outcome measures. Differences in measurement strategies and variable construction further limited our ability to make comparisons across studies, especially for sexual activity and condom use outcomes. After limiting our focus to measures defined relatively consistently across studies, we had a maximum sample of 19 impact estimates for sexual initiation, 24 for recent sexual activity, 11 for condom use at last sex, 14 for condom use consistency, and 17 for sex without a condom. These samples may be too small for drawing definitive conclusions about overarching trends across programs.

The variation in effect sizes may also reflect differences in study population, study setting, and features of the evaluation design. Among the effect sizes reported, some are based on samples of high-risk youth participating in out-of-school or community-based programs, whereas others are based on more general populations of middle- or high-school students. Similarly, whereas some studies report estimates for subgroups of youth (females versus males, for example), others provide report estimates for the full study sample. With a larger number of effect size estimates, we could control for these differences when comparing effect sizes over time or across programs. However, such adjustments would require a greater number of effect size estimates than are currently available.

These findings suggest a clear need for improved standards and reporting of effect size information in teen pregnancy prevention research. When reporting study findings, evaluators should make an effort to include such key information as (a) effect size estimates for both statistically significant and non-significant outcomes; (b) the data used to calculate these effect sizes; (c) a description of effect size calculation methods; and, if possible (d) a discussion of the practical or clinical value of the measured effect sizes and how they compare to external benchmarks or findings from other studies. This expanded reporting will help provide researchers, policymakers, and practitioners with the data needed to inform decision making beyond the preliminary evidence provided in this brief.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

DiClemente, R. J., Wingood, G. M., Harrington, K. F., Lang, D. L., Davies, S. L., Hook III, E. W., ... Robillard, A. (2004). Efficacy of an HIV prevention intervention for African American adolescent girls. *JAMA: The Journal of the American Medical Association*, *292*(2), 171-179.

DiClemente, R.J., Wingood, G.M., Rose, E. S., Sales, J. M., Lang, D. L., Caliendo, A. M.,… Crosby, R. A. (2009). Efficacy of sexual transmitted disease/human immunodeficiency virus sexual risk-reduction intervention for African American adolescent females seeking sexual health services. *Archives of Pediatrics & Adolescent Medicine, 163*(12), 1112-1121.

Downs, J. S., Murray, P. J., Bruine de Bruin, W., Penrose, J., Palmgren, C., & Fischhoff, B. (2004). Interactive video behavioral intervention to reduce adolescent females' STD risk: A randomized controlled trial. *Social Science & Medicine, 59*(8), 1561-1572.

Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics & Adolescent Medicine, 153*(3), 226-234.

Hawkins, J. D., Kosterman, R., Catalano, R. F., Hill, K. G., & Abbott, R. D. (2008). Effects of social development intervention in childhood 15 years later. *Archives of Pediatrics & Adolescent Medicine, 162*(12), 1133-1141.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177.

Jemmott, J. B., Jemmott, L. S., Braverman, P. K., & Fong, G. T. (2005). HIV/STD risk reduction interventions for African American and Latino adolescent girls at an adolescent medicine clinic: A randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine, 159*, (5) 440–449.

Jemmott, J. B., Jemmott, L. S., & Fong, G. T. (1998). Abstinence and safer sex HIV risk-reduction interventions for African American adolescents: A randomized controlled trial. *Journal of the American Medical Association, 279*(19), 1529–1536.

Jemmott, J. B., Jemmott, L. S., & Fong, G. T. (2010). Efficacy of a theory-based abstinence-only intervention over 24 months: A randomized controlled trial with young adolescents. *Archives of Pediatrics & Adolescent Medicine, 164*(2), 152-159.

Jemmott, J.B., Jemmott, L.S., & Fong, G.T. (1992). Reductions in HIV risk-associated behaviors among Black male adolescents: Effects of an AIDS prevention initiative. *American Journal of Public Health, 82*, 372-377.

Jemmott, J.B., Jemmott, L.S., Fong, G.T., & McCaffree, K. (1999). Reducing HIV Risk-Associated Sexual Behavior Among African American Adolescents: Testing the Generality of Intervention Effects. *American Journal of Community Psychology*, *27*(2), 161-187.

Jemmott, J.B., Jemmott, L.S., Fong, G.T., & Morales, K. H. (2010). Effectiveness of an HIV/STD risk-reduction intervention for adolescents when implemented by community-based organizations: A cluster-randomized controlled trial. *American Journal of Public Health, 100*(4), 720-726.

Kirby, D. B., Baumler, E., & Coyle, K. K. (2011). The impact of "Safer Choices" on condom and contraceptive use among sexually experienced students at baseline. (Unpublished manuscript).

Lonczak, H. S., Abbott, R. D., Hawkins, J. D., Kosterman, R., & Catalano, R. F. (2002). Effects of the Seattle Social Development Project on sexual behavior, pregnancy, birth, and sexually transmitted disease outcomes by age 21 years. *Archives of Pediatrics & Adolescent Medicine, 156*(5), 438.

Mindel, C. H. & Hoefer, R. A. (2006). An Evaluation of a Family Strengthening Program for Substance Abuse Offenders. *Journal of Social Service Research, 32*(4): 23-38.

Philliber, S., Williams Kaye, J., Herrling, S., & West, E. (2002). Preventing pregnancy and improving health care access among teenagers: An evaluation of the Children's Aid Society–Carrera Program. *Perspectives on Sexual and Reproductive Health*, *34*(5), 244–251.

Shrier, L. A., Ancheta, R., Goodman, E., Chiou, V. M., Lyden, M. R., & Emans, S. J. (2001). Randomized controlled trial of a safer sex intervention for high-risk adolescent girls. *Archives of Pediatrics & Adolescent Medicine*, *155*(1), 73-9.

Sikkema K. J., Anderson, E. S., Kelly, J. A., Winett, R. A., Gore-Felton, C., Roffman, R. A., Heckman, T. G., Graves, K., Hoffmann, R. G., & Brondino. M. J. (2005). Outcomes of a randomized, controlled community-level HIV prevention intervention for adolescents in low-income housing developments. *AIDS*, *19*(14), 1509–1516.

St. Lawrence, J. S., Brasfield, T. L., Jefferson, K. W., Alleyne, E., O'Bannon, R. E., & Shirley, A. (1995). Cognitive-behavioral intervention to reduce African American adolescents' risk for HIV infection. *Journal of Consulting and Clinical Psychology*, *63*(2), 221–237.

Villarruel, A. M., Jemmott, J. B., & Jemmott, L. S. (2006). A randomized controlled trial testing an HIV prevention intervention for Latino youth. *Archives of Pediatrics & Adolescent Medicine*, *160*, 772-777.

APPENDIX A: REPORT METHODS

Background on the Review. The analysis reported in this brief is part of a broader ongoing systematic review of the teen pregnancy prevention literature conducted for HHS by Mathematica Policy Research and its partner, Child Trends. The review involves four steps: (1) identifying potentially relevant studies for review, (2) screening studies against pre-specified eligibility criteria, (3) assessing each eligible study for methodological quality and risk of bias in the program impact estimates, and (4) identifying programs with evidence of effectiveness in reducing teen pregnancy, STIs, or associated sexual risk behaviors. To date, the review team has identified and assessed about 200 studies released from 1989 through early 2011. From these assessments, the review has identified 31 programs with evidence of a statistically significant impact on at least one measure of sexual activity, contraceptive use, STIs, pregnancy, or births.

Data Collection. To obtain or calculate effect size estimates for the 31 programs highlighted by the review, we reviewed the supporting impact studies and extracted data directly from the reports or journal articles whenever possible. Many studies did not report standardized effect sizes or were missing the information needed to calculate these estimates, such as the analytic sample size or standard deviations (for continuous outcomes). In these cases, we contacted the study authors to request information beyond what they had provided in the original report or journal article.

Calculation Methods. When standardized effect size estimates were not reported directly in the study, we sought to calculate these estimates from other information provided in the study or through our contact with study authors. For the purposes of this brief, we define a "study" as a published report of a program's evaluation results. This means that a single evaluation of a program for which results were reported in multiple publications would be counted as multiple "studies" (e.g. the results from three follow ups of the evaluation of Raising Healthy Children were reported in three separate papers and are counted as three "studies" in this brief). Similarly, a multi-arm evaluation that included more than one of the 31 programs would be counted as multiple "studies" even if the results were reported in a single publication.

We used several decision rules in performing effect size calculations. First, we based our effect size calculations on "point-in-time" differences for all outcomes, regardless of whether the study authors conducted significance testing based on change over time, as in growth curve models. For example, if the study included three follow-up surveys administered at three, six, and twelve months after the program, we calculated an effect size separately for each follow-up, regardless of whether the study's statistical significance tests were calculated separately by follow-up or by pooling data across follow-ups. Second, we calculated different measures of effect size for dichotomous and continuous outcomes. In particular, we calculated standardized mean differences expressed as Cohen's *d* for all continuous outcomes and odds ratios for all dichotomous outcomes. Third, we calculated effect sizes for only those outcomes meeting the review criteria for methodological quality and risk of bias. For example, for quasi-experimental studies, the review criteria require statistical adjustment for a baseline measure of the outcome variable, to minimize the risk of bias in the impact estimates. Our accompanying effect size calculations thus also required regression-adjusted estimates for these designs.

Strengths and Limitations. The approach used for this report has both strengths and limitations. Because we sought to obtain or calculate effect sizes for only the 31 programs highlighted by the review, the findings presented in this brief generalize to only a select number of programs and studies, not the broader range of all teen pregnancy prevention program and studies. In addition, because our approach to calculating effect sizes focused primarily on "point-in-time" estimates (see discussion of calculation methods above), the effect sizes reported in this brief may not always align with the analytic methods featured in the supporting impact study (for example, if the study authors examined change scores or growth curve estimates). Finally, the generalizability of the results is also limited by missing information. We were unable to obtain or calculate effect size estimates for nearly half of the impact estimates considered for the analysis because the necessary information was not provided either in the study or upon follow-up with the study authors.

APPENDIX B: LIST OF PROGRAM NAMES AND ACRONYMS*

Acronym	Program Name	Citation
BART	Becoming a Responsible Teen	St. Lawrence, Brasfield, Jefferson, Alleyne, O'Bannon, & Shirley (1995)
BPBR	Be Proud! Be Responsible!	Jemmott, Jemmott, Fong (1992) Jemmott, Jemmott, Fong, & McCaffree (1999) Jemmott, Jemmott, Fong, & Morales (2010)
CAS	Children's Aid Society (CAS) Carrera Program	Philliber, Williams Kaye, Herrling, & West (2002)
CU	*¡Cuídate!*	Villarruel, Jemmott, & Jemmott (2006)
HZ	Horizons	DiClemente et al. (2009)
MAD	Making a Difference!	Jemmott, Jemmott, & Fong (1998)
MPC	Making Proud Choices!	Jemmott, Jemmott, & Fong (1998)
PHAT-AO	Promoting Health Among Teens! Abstinence-Only Intervention	Jemmott, Jemmott, & Fong (2010)
PHAT-C	Promoting Health Among Teens! Comprehensive Abstinence and Safer Sex Intervention	Jemmott, Jemmott, & Fong (2010)
RHC	Raising Healthy Children	Hawkins, Catalano, Kosterman, Abbott, & Hill (1999) Lonczak, Abbott, Hawkins, Kosterman, & Catalano (2002) Hawkins, Kosterman, Catalano, Hill, & Abbott (2008)
SC	Safer Choices	Kirby, Baumler, & Coyle (2011)
SiHLE	Sisters Informing, Healing, Living, and Empowering	DiClemente et al. (2004)
SS	Safer Sex	Shrier et al. (2001)
SSS	Sisters Saving Sisters	Jemmott, Jemmott, Braverman, & Fong (2005)
TOP	Teen Outreach Program	Allen, Philliber, Herrling, & Kuperminc (1997)
WCYD	What Could You Do? / 17 Days	Downs et al. (2004)

*** Note:** Effect sizes from a number of studies were obtained or estimated but not reported after limiting our focus to measures defined relatively consistently across studies. Effect sizes for the following 12 evidence-based programs are not reported in this Brief, but are available online at http://tppevidencereview.aspe.hhs.gov: Aban Aya Youth Project, Adult Identity Mentoring (Project AIM), Assisting in Rehabilitating Kids, Be Proud! Be Responsible! Be Protective!, Draw the Line/Respect the Line, FOCUS, Heritage Keepers Abstinence Education, It's Your Game: Keep It Real, Project TALC, Rikers Health Advocacy Program, Sexual Health and Adolescent Risk Prevention, and Teen Health Project.

DEPARTMENT OF HEALTH
& HUMAN SERVICES

Office of the Secretary
Washington, DC

OFFICIAL BUSINESS
Penalty for Private Use $300