

Teen Pregnancy Prevention Replication Study: Short-Term Impacts of *Reducing the Risk*

|EVALUATION REPORT|

OCTOBER 2016

Reducing the Risk: Short-Term Impact Report

Teen Pregnancy Prevention Replication Study

October 2016

Prepared for:
Amy Farb
The Office of Adolescent Health

Lisa Trivits
Sarah Oberlander
The Office of the Assistant Secretary for Planning and Evaluation

Submitted by:
Abt Associates
55 Wheeler St.
Cambridge, MA 02138

In Association with:
Decision Information Resources, Inc.
and
Belmont Research Associates

Recommended Citation

Kelsey, M., Layzer, J., Price, C., Juras, R., and Blocklin, M. “Reducing the Risk Short-Term Impact Report: Findings from the Teen Pregnancy Prevention Replication Study”. Cambridge, MA: Abt Associates Inc., October 2016.

This document is available at https://aspe.hhs.gov/teen-pregnancy-prevention-tpp-replication-study

This document was prepared under contract number HHSP23320095624WC Order No. HHSP23337011T (awarded in September 2011) from the U.S. Department of Health and Human Services, Office of Adolescent Health and the Office of the Assistant Secretary for Planning and Evaluation. Any statements expressed are those of the authors and do not necessarily reflect the views of the Office of Adolescent Health, the Office of the Assistant Secretary for Planning and Evaluation, or the U.S. Department of Health and Human Services.

Acknowledgements

We would like to thank many individuals and organizations who helped to make this evaluation possible. First, we thank the staff from the Office of Adolescent Health (OAH) within the U.S. Department of Health and Human Services (HHS) for their support throughout the project. In particular Amy Farb, the OAH project officer for the Teen Pregnancy Prevention Replication Study, provided critical guidance throughout the project. We also thank Amy Margolis, the many program officers and Alexandra Warner. Lisa Trivits and Sarah Oberlander, the project officers for the TPP Replication Study from the Office of the Assistant Secretary for Planning and Evaluation (ASPE) provided critical direction and support throughout the project and helped to ensure the success of the study.

We would like to thank our project partners Decision Information Resources (DIR) for their commitment and hard work in support of the study and data collection efforts, along with Innovative Clinical Research Solutions (ICRS). Belmont Research Associates provided invaluable leadership and project direction, and helped navigate many obstacles and establish critical relationships on the ground. A special thanks to SANDAG and Health Advocates for local evaluation support. We would like to acknowledge each of the grantee sites whose close partnership and commitment were essential to the success of the study. Project staff not only implemented the program, but participated in interviews and remained dedicated to the study throughout.

At Abt, the project was supported by a team of individuals who worked closely with each of the grantee sites. Kim Francis and Lesley Freiman provided essential support for implementation of the study. Rob Olsen provided guidance and support during the initial study design phase, and Fatih Unlu provided valuable feedback on the report. Thanks to Dan Weiss and Heather He for programming and technical assistance, and Stefanie Falzone and Jan Nicholson for producing the report.

And finally, a tremendous thanks to all of the youth who participated in the study. Without them, this entire endeavor would not have been possible.

Contents

Exhibits

1. Introduction

Reducing rates of unplanned teen pregnancy and sexually transmitted infections (STIs) are priorities for the U.S. Department of Health and Human Services (DHHS). To achieve this goal, the Department is investing in evidence-based pregnancy reduction strategies and targeting populations at highest risk for teen pregnancy. The federal Teen Pregnancy Prevention (TPP) Program, administered by the Office of Adolescent Health (OAH), includes funding for programs that are intended to address high rates of teenage pregnancy by (1) replicating evidence-based models, and (2) testing innovative strategies.

The TPP Program was authorized in 2010 as part of the larger Teen Pregnancy Prevention Initiative. The program initially included $100 million in annual funding to support programming. Of these funds, $75 million were available annually to support five-year grants for replicating 28 program models that prior rigorous evaluations had shown to be effective. These program models were identified through a systematic, comprehensive review of the literature on teen pregnancy, STIs, and sexual risk behaviors (Kappeler & Farb, 2014).

The TPP Program also acknowledges the limitations of existing research and the need for additional research on programs, citing lessons learned from the comprehensive evidence review such as an absence of independent evaluations and a limited number of program replications (Goesling et al., 2014). In the review, evidence for many of the programs rests on a single study of effectiveness, often conducted a long time ago and with a single population. A program may work in one location with a particular population, but that does not necessarily mean it will be effective in another. Further, implementing a program model with fidelity often competes with the need to adapt to local conditions on the ground. For these reasons, a carefully designed study of multiple replications of selected program models is an important contribution to the existing research.

1.1 The Replication Study

The TPP Replication Study[1] is being conducted for OAH, under a contract with the Office of the Assistant Secretary for Planning and Evaluation (ASPE), by Abt Associates and its subcontractors, Belmont Research Associates, Decision Information Resources (DIR), and CiviCore. The study has two major components: an Impact Study and an Implementation Study.

Impact Study. Through a series of rigorous experimental design evaluations, the study tests multiple replications of three evidence-based program models to determine their effectiveness across different settings and populations. The strategy of selecting multiple replications of each program model allows us to pool data across the three replication sites to assess impacts on such behavioral outcomes as pregnancy and for key subgroups (e.g., those based on age and sexual experience). In addition, the strategy lets us examine variation in impacts across replications for each program model and gather evidence about the generalizability of program effectiveness.

Implementation Study. A comprehensive Implementation Study will provide information about the contexts in which the evidence-based programs were implemented and the challenges faced in implementing them. It will also allow us to assess aspects of program implementation that are associated with program impacts.

[1] The study is also referred to as the Teen Health Empowerment Study in the field with program staff and study participants.

1.2 The Three Models Replicated

OAH, with its ASPE partners, selected three program models from the first round of TPP-funded grants to test and replicate: the *Safer Sex Intervention* (a clinic-based HIV/STI prevention program for high-risk adolescent females); *Reducing the Risk* (a sexual health education curriculum); and *¡Cuídate!* (an HIV/STI risk reduction program targeting Latino youth). These programs were selected based on the breadth and scale of the proposed replication effort. All three were proposed for replication by at least five grantees.[2] In addition, the three program models represent a range of targeting and service strategies, as well as some variation in the settings in which services are provided.

1.3 Focus of This Report

This report, which focuses on the short-term impacts of *Reducing the Risk (RtR)*, is one in a series of reports that present findings on the implementation and effectiveness of the three program models. This report presents findings from the first of two follow-up surveys designed to examine the short-term and longer-term impacts of *RtR*. Two companion reports examine the short-term impacts (six to 12 months post-baseline[3]) of the remaining program models in the study. Three final impact reports will present findings on the longer-term impacts (18 to 24 months post-baseline) of all three program models. Three Implementation Study reports will document the implementation of each of the program models. In addition, nine site profiles provide an overview of program implementation as well as descriptive information about the study participants at baseline in each site.[4]

[2] Of the 28 program models in the TPP Program, the *Teen Outreach Program (TOP)* is the most frequently replicated. Seven independent evaluations of TOP were conducted as a condition of those grants. For this reason, it was excluded from consideration for the TPP Replication Study. *Becoming a Responsible Teen (BART)*, another widely used model, was also excluded because it had already undergone several evaluations.

[3] Where "baseline" means the point at which each study participant entered the study.

[4] The profiles are available at https://aspe.hhs.gov/basic-report/tpp-replication-study.

2. The Program Model: *Reducing the Risk*

RtR is a sexual health curriculum developed in the early 1990s to help prevent pregnancy and STI transmission in high school–age adolescents. The curriculum focuses on changing four sexual behaviors directly related to this goal: initiation of sexual intercourse, abstinence, use of condoms, and use of contraception. *RtR* is intended for use in high school classrooms with students of all ethnicities, although program materials suggest it can be delivered in community settings. *RtR* consists of 16 modules of 45 minutes each. The modules can be delivered separately or grouped into eight 90-minute sessions, but must be delivered in their specified sequence.

The program's objectives for student participants are that they will be able to:

- Evaluate the risks and consequences of becoming a teen parent or becoming infected with an STI;
- Recognize that abstinence and the use of contraception are the only ways to avoid pregnancy;
- Conclude that factual information is essential to avoid pregnancy or STIs; and
- Demonstrate effective refusal and negotiation skills. (Lezin et al., 2010)

The three behavioral theories that provide the basis for *RtR* all hypothesize that, to reduce or avoid risky behavior, people need to learn and personalize relevant information, recognize social pressures and anticipate risky situations, establish norms for positive behaviors, and learn and practice skills so that they can act on the information (Lezin et al., 2010). Accordingly, although the *RtR* program includes mini-lectures and worksheets, it places great emphasis on skills practice and problem solving through group discussions and role play.

2.1 Content and Pedagogy

RtR is a highly scripted program in which core content and pedagogical elements are specified in detail, as is the module in which they should be covered (or used, in the case of pedagogical strategies). Exhibit 2.1 shows these curricular components mapped by module. The program also dictates that the trained teachers or health educators delivering it be comfortable discussing sexuality, model skills during role play, give clear directions, and tailor the language they use to connect better with the youth served.

Exhibit 2.1: *Reducing the Risk* Core Content and Pedagogical Components by Module

Core content component	Module
Knowledge about:	
Pregnancy risk	1
HIV and other STI prevention, transmission, treatment, and consequences	1A,[5] 12
Abstinence	1, 2
Birth control methods and effectiveness	7, 8
How to access health care information and contraceptives	7, 8
Elements of successful relationships	2

[5] Implementers of the program can choose either Module 1 or Module 1A or, if there is time in the schedule, deliver both modules to strengthen the message about STI risk delivered in a later module.

Core content component	Module
Effective refusal skills and delaying tactics	3, 6
Attitudes about:	
Abstinence	2
Having sex and unprotected sex	3
Using condoms and other contraception	14
HIV risk and consequences	12
Pregnancy risk and consequences	1
Skills and self-efficacy to:	
Refuse sex and unprotected sex	4
Delay sex	16
Use refusal, delay, and communication in pressure situations	10
Obtain information and condoms/contraception	7
Negotiation to use condoms/contraceptives	11
Perception of risk of:	
Pregnancy	1
HIV	13
Being in unprotected "risk crisis"	6
Social/peer norms about:	
Sex and abstinence	2, 15
Condom use	9, 15
Values:	
Understanding parent/adult values about teen sexual activity	3
Intentions to:	
Use refusal skills and delaying tactics	5
Be abstinent	15
Use condoms/contraception	14
Avoid pregnancy	8
Avoid HIV	12
Communication:	
With parents/other adults about teen sexual activity	3, 6
Create a learning environment by:	
Providing a well-thought-out introduction	1
Setting ground rules	1
Summarizing previous lesson	1–16
Reviewing current lesson	1–16
Following detailed steps for each activity	1–16

Core content component	Module
Facilitate learning activities by using:	
Repetition to reinforce learning	Throughout
Lectures	1, 1A, 2, 3, 5, 6, 7, 8
Role plays	1, 1A, 3, 4, 5, 9, 10, 11, 14, 16
Large-group discussion	1A, 2, 3, 15, 16
Brainstorming	2, 6, 8, 15
Guest speakers (as alternative to clinic visit)	8
Worksheets	1, 1A, 5, 6, 9, 10, 11, 14, 15, 16
Homework followed by large-group discussion	3, 7, 8, 12
Traffic light exercises	13
Quizzes	4, 5
Address multiple learning styles by:	
Using a variety of teaching methods	Throughout

2.2 *Reducing the Risk* Logic Model

RtR is probably the earliest of the comprehensive sexual health programs, and it provided a basis for many later models. It is widely used across the United States, although it often is used in an abbreviated form when schools are unwilling or unable to accommodate its 16 modules into their schedules.

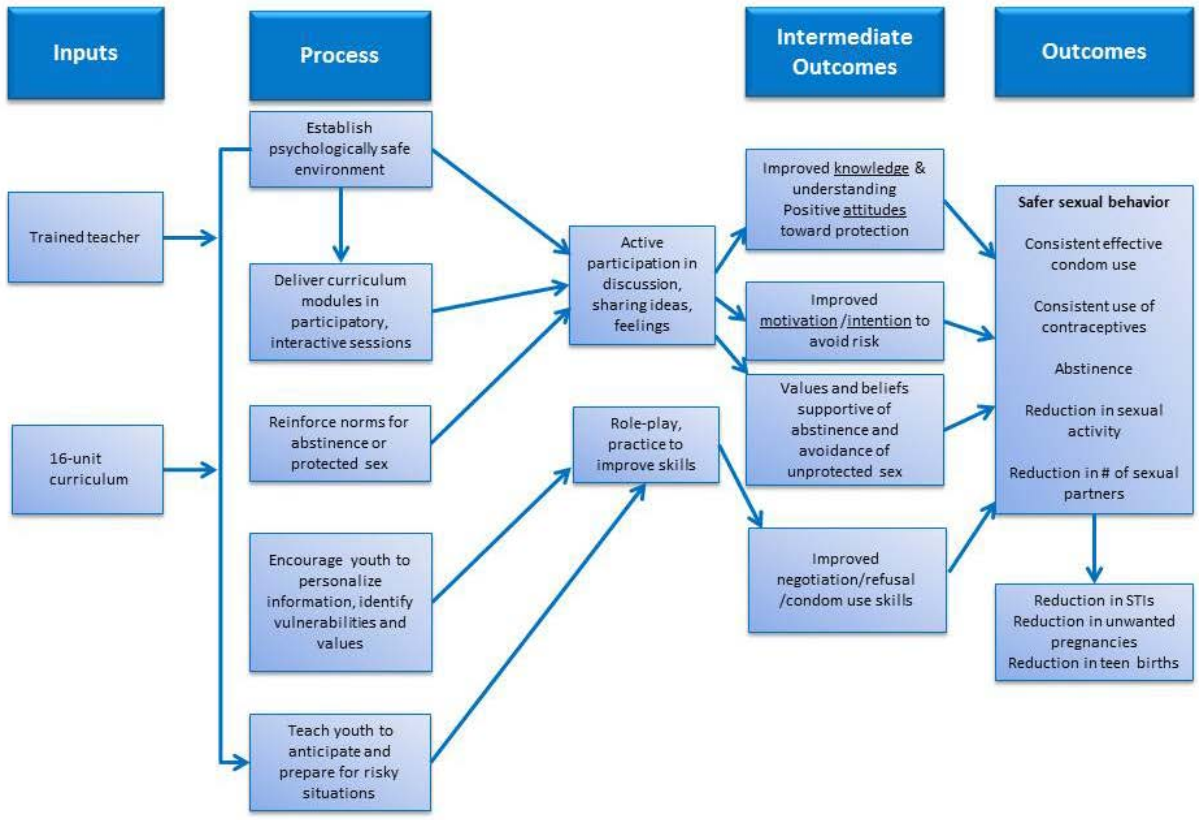
Exhibit 2.2 (on the next page) shows the program elements, the intended outcomes, and the pathways by which the program seeks to achieve these outcomes. The teacher or health educator delivers *RtR* in a classroom or other setting. The first objective for the teacher is to create a psychologically safe environment of mutual trust in which youth can speak freely about their attitudes, feelings, values and perceptions. Within that atmosphere of trust, the teacher delivers the 16 modules in a prescribed sequence. As part of every module, the teacher reinforces the norms of abstinence and protected sex.

The sessions are interactive and encourage active participation by students. Youth are encouraged to personalize the information, identify their own vulnerabilities, and examine their personal values. The modules repeatedly offer opportunities for youth to anticipate and prepare for situations in which they may be pressured to have unwanted or unsafe sex, and to practice the skills they will need to deal with these and similar situations.

Taken together, the program's modules are intended to increase students' knowledge and understanding of sexual health issues; improve students' attitudes toward protection; improve motivation and intentions to avoid risk; encourage values and beliefs that are supportive of abstinence and avoidance of unprotected sex; and improve negotiation, refusal and condom use skills.

These intermediate outcomes are expected to lead to the behavioral outcomes that the program seeks to achieve: correct and consistent use of condoms and birth control for those youth who are sexually active; abstinence from sex; and reductions in sexual activity and number of partners. Avoidance of or reduction in sexually risky behavior is expected to reduce rates of STIs, unwanted pregnancies, and births among teens.

Exhibit 2.2: *Reducing the Risk* Logic Model



Each replication adhered to the core components of the model, but modifications were made to comply with state mandates, to address gaps in program content, or to accommodate local constraints (see Section 3.2.2).

2.3 Evidence of Effectiveness

RtR is one of the programs identified as evidence-based by the HHS Teen Pregnancy Prevention Evidence Review that TPP grantees could choose to implement (HHS, 2010). As with many other program models identified through this review, evidence for its effectiveness comes from a single study that was completed many years ago (25 years ago, in the case of *RtR*).

The evaluation of *RtR* employed a quasi-experimental design and was conducted in 46 classes in rural and urban schools in northern California (Kirby et al., 1991). It compared students in *RtR* classrooms with students in classrooms where a standard health class was taught. Almost two-thirds of the 758 students who participated in the study were White, 20 percent were Hispanic, and 2 percent were Black. A little more than half of the students were 10th graders, one-quarter were 9th graders, and the remainder were 11th and 12th graders.

The evaluation conducted surveys 6 and 18 months after the intervention ended (Kirby et al., 1991). Analysis of the survey data found the following:

- There were no impacts on recent sexual activity at 6 months.[6] However, at 18 months, among students who were sexually inexperienced at baseline, significantly fewer *RtR* participants had initiated sexual intercourse than had students in the comparison group.
- There were no effects on unprotected sex at 6 months. However, at 18 months, among students who were sexually inexperienced at baseline, *RtR* participants were significantly less likely to have had unprotected sex than students in the comparison group. Among these sexually inexperienced students, females and lower-risk *RtR* participants were significantly less likely to report having had unprotected sex than were their counterparts in the comparison group.
- *RtR* had significant positive impacts on participants' knowledge of pregnancy and STI risk and on their perceptions of the proportion of their peers who had ever had sexual intercourse.

[6] "Recent" was defined as in the last 90 days.

3. Evaluation Design

The Impact Study is designed to estimate the effects of three replications of *RtR*.[7] It addresses questions about the effects of the program on participants' sexual behaviors, as well as on the intermediate outcomes the logic model predicts will lead to the behavioral outcomes that *RtR* seeks to achieve.

The current report focuses on short-term program effects 12 months after study enrollment. It is guided by the research questions below.[8]

3.1 Research Questions

- *Did* RtR *increase teens' exposure to information on reproductive health, contraception, and STI transmission and prevention?*
- *Did* RtR *improve teens' knowledge and understanding of reproductive health, risky sexual behavior, pregnancy prevention, and the transmission and prevention of STIs?*
- *Did* RtR *have positive effects on teens' attitudes toward sexual activity, birth control, and condom use?*
- *Did* RtR *increase teens' motivation to delay childbearing?*
- *Did* RtR *decrease teens' intentions to engage in risky sexual behavior?*
- *Did* RtR *increase teens' confidence in their ability to refuse unwanted sex and/or to negotiate safe sex?*
- *Did* RtR *delay initiation of or reduce teens' risky sexual behavior?*
- *Do program impacts differ by replication site and for key subgroups (e.g., gender, age, race/ethnicity, sexual experience at baseline)?*

3.2 Study Design

In each of the replication sites, the study employed an experimental design in which classes were randomly assigned to receive the *RtR* intervention or to a control group that received the usual curriculum (e.g., physical education (PE), science, health, social studies, or other non-core subject). This section describes the selection of the three replication grantees, site-specific program designs, settings for the program, recruitment and random assignment, and the treatment and control conditions.

3.2.1 Selection of Replication Grantees

The study design called for evaluating at least three replications of the model. At the time of site selection for the study, *RtR* was being replicated by at least five grantees. Complicating the selection was the fact

[7] A more detailed impact study design report can be found at https://aspe.hhs.gov/basic-report/tpp-replication-study.

[8] The final impact report will answer a similar set of questions about program effects on intermediate outcomes and risk behaviors after two years. It also will examine program impacts on pregnancy.

that most grantees had not planned for a rigorous evaluation.[9] In some cases, schools, districts, or other partners had signed agreements with grantees to implement the *RtR* program but had no such agreements about evaluation. Sometimes these agreements could be renegotiated. In other cases, districts were unwilling to honor the agreements if that meant participating in a rigorous evaluation. In still other cases, grantees were struggling to reach agreements with school districts to implement the program, and it was unclear whether they would be successful with the added burden of an evaluation. These considerations led us to eliminate some potential candidates.

The three grantees selected were:

- **Better Family Life**. Better Family Life (BFL) is a nonprofit community development agency with deep roots in the St. Louis, Missouri, metropolitan area. Established more than 30 years ago, BFL partners with more than 50 organizations in the region to provide services to more than 50,000 individuals, most of whom are low-income and Black. Although youth workforce development is a major focus, the agency manages a variety of after-school programs in multiple school districts. In 2004, BFL moved to address sexual health issues and the skills needed to build healthy relationships, delivering services in schools and community-based organizations across the St. Louis metropolitan area.
- **LifeWorks.** LifeWorks is a private nonprofit agency that offers housing, counseling, education, workforce, and youth development programs to more than 6,000 youth and their families in locations across Travis County, Texas. Since 1997, the agency has provided teen pregnancy prevention education and support services to middle school youth and preadolescents in Travis County. For this population, the agency offers programs that focus on strategies to resist peer pressure, build self-esteem, delay sexual activity, and make healthy choices. For the TPP grant, LifeWorks partnered with Planned Parenthood of Greater Texas to deliver *RtR.*
- **San Diego Youth Services.** San Diego Youth Services (SDYS) is a nonprofit agency that provides services to help young people who are at risk for not achieving self-sufficiency. SDYS provides a spectrum of services including housing, family-centered counseling, and life-skills training for at-risk youth; individual counseling for youth recovering from addiction; and after-school programs, to more than 13,000 youth and families annually at 14 locations across San Diego County, California. For the TPP grant, SDYS partnered with four other multi-service agencies that serve youth and families in different areas of the county. Together, the five agencies cover all of what is a very large county, offering services in all of its 18 cities and implementing *RtR* in schools countywide.

3.2.2 Site-Specific Program Designs

In all three replication sites, grantees proposed to deliver all 16 modules of the *RtR* program.

[9] The 2010 TPP grant program offered multiple funding ranges. All funded projects were expected to monitor and report on program implementation and outcomes through performance measures. Projects in the higher funding range (greater than $1 million per year) were expected to be implemented in multiple sites within a targeted geographic area and were required to undertake an independent local evaluation. Projects in the lower funding range (less than $1 million per year) were not expected to undertake a rigorous local evaluation. Two of the *RtR* replications selected for the study were in this lower range. San Diego Youth Services (SDYS), a larger-scale replication, had proposed a rigorous local evaluation.

In its grant proposal, BFL requested and received permission from OAH for one school to deliver the program to male and female students separately, with a health educator of the same gender. To respond to concerns in the same school, BFL was allowed to replace the condom demonstration with a video (for boys) and a mini-lecture (for girls). LifeWorks received permission to add two additional modules: one on reproductive anatomy and the optional *RtR* introductory module, reinforcing messages about pregnancy and STI prevention. LifeWorks was also permitted to drop *RtR*'s condom demonstration, to conform to district policy. SDYS received permission to deliver the program with two instructors, to accommodate larger class sizes.

3.2.3 Settings for the Program

Grantees were selected for participation in the study between six and seven months after they had received the grant award. This meant that, for the most part, the grantees had recruited implementation locations prior to being selected for the study. In all three replications, *RtR* was delivered in public school classrooms, as part of the regular school day.

BFL delivered the program in 9th grade classes (with a small number of students from higher grades) in six public high schools in St. Louis City and St. Louis County in Missouri, and in St. Clair County, Illinois. The LifeWorks replication implemented the program in health classes (mixed 9th and 10th grades, with some older students) in five public high schools in the Austin Independent School District. SDYS and its four partners implemented the program in 8th or 9th grade PE, health, or science classes in six public middle, junior high, and high schools in San Diego County.

3.2.4 Recruitment and Random Assignment

In each replication site, once schools had agreed to participate, school staff identified the classes that would participate in the study (and be randomly assigned to *RtR* or to the usual curriculum). Across the three replication sites, school staff selected a total of 150 classes for the study. At the beginning of the semester, class rosters for the selected classes were provided to the Abt study team, and students were recruited for the study in each of the classes identified (without knowledge of whether a class would receive *RtR* or the usual curriculum).

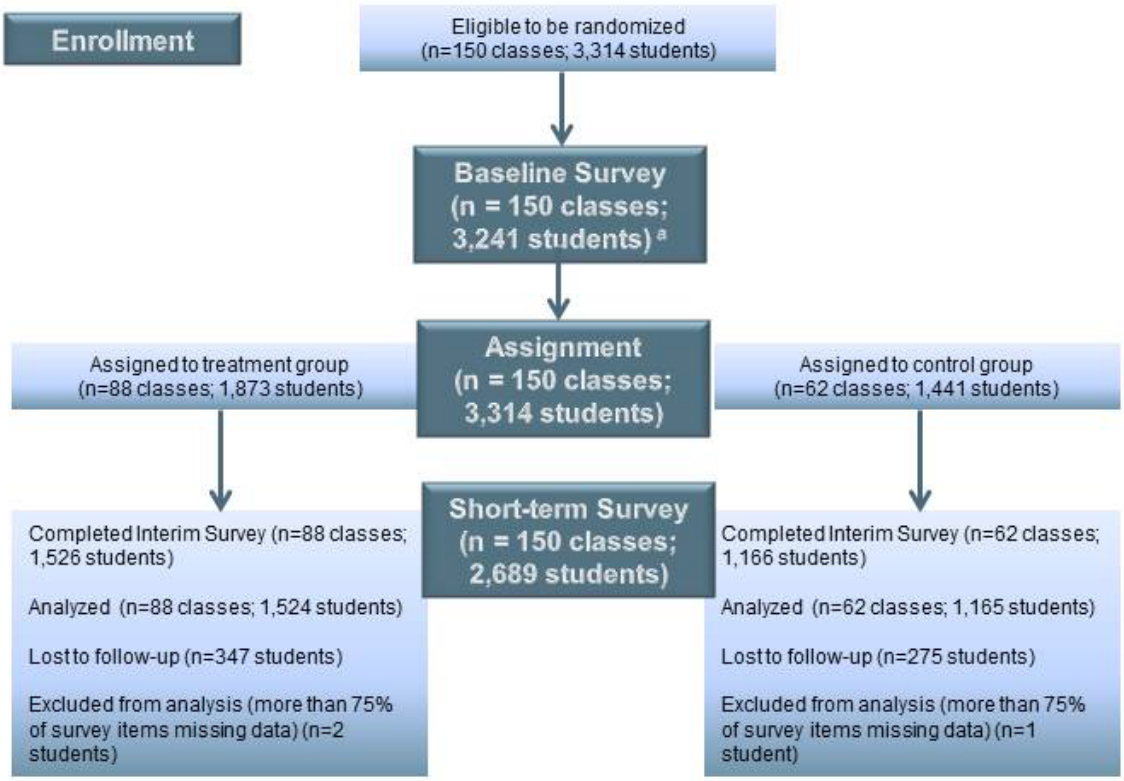
Grantee and partner staff who had been carefully trained by Abt study staff conducted presentations to each class. These presentations included information about the study procedures, a practical illustration of random assignment, and a description of the treatment and control conditions. The presentations were intended to personalize the study and help in recruiting students. Grantee and partner staff distributed parental consent forms and study brochures and provided teachers with small incentives for the return of signed forms. They worked with individual teachers to obtain parental consent, and notified the Abt team about students whose parents had consented to their children participating in the study. Once the period allowed for consent had expired for a given school, a baseline survey was administered to all students participating in the study.

Abt staff randomly assigned classes within each school to either the treatment group (*RtR* to be delivered) or the control group (usual curriculum to be delivered). The random assignment ratio varied across replication sites and schools, based on school and program preferences, with more classes assigned to treatment overall. Across the 150 classes, 88 classes (1,873 students) were assigned to the treatment group, and 62 classes (1,441 students) were assigned to the control group.

Once students had completed the baseline survey, teachers, schools, and students were informed of the random assignment results. Students whose parents had not given consent were reassigned to a different class scheduled for the same class period.

Exhibit 3.1 shows how we arrived at the study's analytic sample via random assignment and the survey completion process, starting with the 3,314 eligible students (i.e. those in the 150 classes selected to be randomized) who obtained parent permission.

Exhibit 3.1: Study Sample



[a] A total of 73 participants did not take the baseline survey. Among them, 43 were assigned to the treatment group and 30 were assigned to the control group.

3.2.5 Treatment and Control Conditions

Across the three replications, health educators[10] hired by the grantee or a partner agency delivered all 16 modules of *RtR*. (LifeWorks' staff also delivered the additional two sessions described earlier.) Health educators were trained by the program distributor, and supervised and monitored continuously by supervisory staff in each agency and local evaluators. Fidelity, quality, and attendance measures required by OAH were completed and reported by health educators, supervisory staff, and external evaluators.

[10] The staff who delivered *RtR* in the schools were called health educators or facilitators. We have used "health educators" as a blanket term.

Beyond these commonalities there were minor differences in the schedules for delivering the curriculum, in the age range of student participants, and in the size of the classes. In one of the replication sites, members of the control group all participated in the same standardized activity (the scheduled health class). In the other two replications, control group members attended the regularly scheduled class, which differed by school (Exhibit 3.2).

Exhibit 3.2: Treatment and Control Conditions in the Three Replications

Grantee	Treatment Group	Control Group
Better Family Life	Number of modules: 16 Delivery schedule: Delivered in sixteen 45-minute classes Gender and size of groups: Average-size, mixed-gender classes (one school held separate classes for male and female students, with instructors of matching gender)	PE, ROTC, health, homeroom/guidance classes at schools' discretion
LifeWorks	Number of modules:18 Delivery schedule: Delivered in nine 90-minute sessions (health class) Gender and size of groups: Average-size, mixed-gender classes	Health class
San Diego Youth Services	Number of modules:16 Delivery schedule: Delivered in eight 90-minute or sixteen 45-minute classes Gender and size of groups: Large, mixed-gender or same-gender groups, depending on school, with two health educators	PE, health, or science classes at schools' discretion

3.3 Measures and Data Collection Strategies

3.3.1 Data Collection Strategy

To assess the impacts of the *RtR* intervention, participating students in the three replication sites were surveyed three times: at baseline, before the program began; 12 months after the baseline survey (short-term follow-up); and 24 months after the baseline survey (longer-term follow-up). This report assesses outcomes using data from the 12 month survey; findings from the 24-month survey will follow in a later report. For each survey, a web-based Audio Computer-Assisted Self-Interview (ACASI) system was used to capture and store survey responses, and respondents could choose to take the survey in Spanish or English. At baseline, paper copies of the survey (in Spanish and English) were available as backup in case of computer or Internet failure.

The baseline survey was completed in a group setting at the school. Individual students used school computers, where possible, or tablets dedicated to the study, if not. Study staff oversaw the baseline survey and distributed gift cards to students upon completion. Make-up survey days were arranged, to allow as many participants as possible to complete the survey. Of the 3,314 eligible students, 3,241 (97.8 percent) completed a baseline survey.

For the first follow-up survey that is the subject of this report (12 months after baseline), only the web-based ACASI system was used. For tracking purposes and to invite/remind students to complete their survey, students were sent emails and texts before the survey went live and throughout the survey

period[11]. In some cases, before the survey period closed, field staff contacted participants and encouraged them to complete the survey independently online or helped them to access the survey. Participants could complete the survey using personal tablets or computers, school or library computers, or even their smart phones. Gift cards were mailed to participants after completion.

As Exhibit 3.3 shows, 81 percent of eligible students completed the short-term follow-up survey (12 months post-baseline). There was almost no difference in the response rates of students in the treatment group versus those in the control group. Of the three replication sites, BFL had the highest response rates.

Exhibit 3.3 *Reducing the Risk* Short-Term Impact Survey Response Rate

	Participants			Completed Short-Term Follow-Up Survey					
				Total		Treatment		Control	
	Total	Treatment	Control	N	Percent	N	Percent	N	Percent
All Sites	3,314	1,873	1,441	2,689	81.1%	1,524	81.4%	1,165	80.9%
Better Family Life	1,050	640	410	941	89.6%	572	89.4%	369	90.0%
LifeWorks	1,093	568	525	853	78.0%	442	77.8%	411	78.3%
San Diego Youth Services	1,171	665	506	895	76.4%	510	76.7%	385	76.1%

3.3.2 Measures

The first follow-up survey collected information from students on a variety of factors, including questions that allow us to measure three sets of outcomes: (1) exposure to information about topics related to sexual risk behavior; (2) intermediate outcomes (i.e., factors that the *RtR* model predicts will lead to behavioral outcomes); and (3) sexual activity and risk behavior. We briefly describe these measures here. A more complete description of these measures and the individual survey items that comprise them can be found in Appendix C. Exhibit 3.4 (on the next page) summarizes the outcome measures and their construction.

Exposure to Sexual Health Information. In the first follow-up survey, we asked students about their exposure to information about reproductive health and related topics. Students were asked whether they had received information about any of a set of topics in the 12 months preceding the survey. Because the topics were distinct, we examined responses to individual survey questions, rather than creating and analyzing a composite measure.

Intermediate Outcomes. Drawing on knowledge of the program's theory of change and exploratory factor analysis, we constructed composite measures to assess four factors that potentially lead to behavioral outcomes: (1) ***knowledge*** *of pregnancy risk* and *knowledge of STI risk*; (2) ***attitudes*** *toward protection* and *attitudes toward risky sexual behavior*; (3) ***motivation*** *to delay childbearing*; and (4) *condom negotiation and refusal* ***skills***. For a fifth measure, we analyzed four single-item measures: (5) ***intentions*** to become sexually active in the immediate future and to use protection when sexually active.

> ***Knowledge*****.** We constructed two composite measures: *knowledge of pregnancy risk* and *knowledge of STI risk.* The 4 items that make up the first measure and the 12 items that make up the second are all factual questions, testing students' knowledge of the circumstances under which a woman can become pregnant (e.g., "A woman is protected from pregnancy the day she begins taking the pill")

[11] Participants could complete the survey within a three month window.

and the effectiveness of condoms and other methods of contraception (e.g., "If birth control pills are used correctly and consistently, how much can they decrease the risk of pregnancy?"), as well as facts about STIs and their transmission (e.g., "You can't get infected with HIV if you have sex only once or twice without a condom"). All items were scored 1 for a correct answer and 0 for any incorrect answer; scores were averaged across the items that make up a measure and multiplied by 100 to indicate the percentage of items answered correctly.

Attitudes. We constructed two composite measures of attitudes: *attitudes toward protection* and *attitudes toward risky sexual behavior*. For the 12 items that make up the first measure (attitudes toward protection), students were asked if they agreed or disagreed with statements such as "Birth control is important to make sex safer." Four response categories ranging from "strongly disagree" to "strongly agree" were scored from 1 to 4, and then scores for individual items were averaged, with higher values representing more positive attitudes toward the use of protection. For the seven items that comprise the second measure (attitudes toward risky sexual behavior), students were asked whether they agreed or disagreed with statements such as "It's OK to have sex with someone on the first night you meet them." Responses were scored 0 (disagree) or 1 (agree), averaged across the items, and multiplied by 100 to indicate the percent of items agreed with, with higher scores representing higher levels of support for risky behavior.

Motivation. We constructed one composite measure, *motivation to delay childbearing*, which was made up of three survey items that asked respondents whether they agreed or disagreed with statements such as "It is important for you to finish school before you have a child." Four response categories ranging from "strongly disagree" to "strongly agree" were scored from 1 to 4. Scores for individual items were averaged, with higher values representing greater motivation to delay childbearing.

Skills. We constructed two composite measures of skills: *condom negotiation skills* and *refusal skills.* The measure of *condom negotiation skills* was made up of seven items asking about respondents' perceptions of their ability to obtain and negotiate the use of condoms with a partner (e.g., "If you were going to have sex, could you insist on using a condom even if your partner didn't want to use one?"). Possible responses ranged from "I'm sure I could not" to "I'm sure I could," coded 1 to 4. All items were averaged, with higher scores representing greater certainty of condom negotiation skills. The measure of *refusal skills* comprised six items probing respondents' perceptions of their ability to refuse to engage in sexually risky behavior (e.g., "How sure are you that you would be able to say no to having sexual intercourse if neither you nor your partner had any form of birth control?"). Possible responses ranged from "I'm sure I could not" to "I'm sure I could" and were coded 1 to 4. Scores for individual items were averaged, with higher scores representing greater certainty of refusal skills.

Intentions. We included four single-item measures of students' intentions related to sexual activity in the year following the survey. The first item asked about oral sex; the second about sexual intercourse; the third about condom use; and the fourth about contraception use. Responses to each of the four items were scored 0 or 1 (1 for those responding that their intentions were "probably" or "definitely").

Sexual Activity and Sexual Risk Behavior. To address the study's most important questions about the impact of the intervention, we identified seven measures in the domain of youth sexual behavior at the short-term follow-up: (1) *currently sexually active (in the last 90 days), ,* (2) *sexual intercourse in the last 90 days,* (3) *oral sex in the last 90 days,* (4) *initiation of sexual activity,* (5) *sexual intercourse without*

birth control (in the last 90 days), (6) *sexual intercourse without a condom (in the last 90 days),* and (7) *oral sex without a condom (in the last 90 days).*

Sexual activity is defined differently across replication sites. In BFL, "sexual activity" refers to sexual intercourse, oral sex, and/or anal sex. Students were not asked about anal sex in LifeWorks or SDYS: in these sites, "sexual activity" refers to sexual intercourse or oral sex. All seven measures are single items, with yes/no answers.

Exhibit 3.4: Outcome Measures

Measure	Definition
EXPOSURE TO SEXUAL HEALTH INFORMATION	
Exposure to sexual health information	Eight single items reflecting exposure to information about: (a) relationships or marriage; (b) abstinence; (c) birth control methods; (d) where to obtain birth control; (e) STIs; (f) how to talk with a partner about sex and birth control; (g) how to say no to sex; and (h) how babies are made. Responses were coded as 1 = yes and 0 = no.
INTERMEDIATE OUTCOMES	
Domain: Knowledge	
Knowledge of pregnancy risk	Continuous index: average of responses to four questions about circumstances in which it is possible to become pregnant and the extent to which contraceptive methods protects against pregnancy. Average scores multiplied by 100 range from 0 to 100 and represent the percentage of the four questions answered correctly, with higher values representing more accurate knowledge.
Knowledge of STI risk	Continuous index: average of responses to 12 questions about STI transmission and prevention multiplied by 100. Scores range from 0 to 100 and represent the percentage of the 12 questions answered correctly, with higher values representing more accurate knowledge.
Domain: Attitudes	
Attitudes toward protection	Continuous index: average of responses to 12 questions about attitudes toward using condoms and/or birth control during sex. Average scores range from 1 to 4 with higher values representing more positive attitudes toward using protection.
Attitudes toward risky sexual behavior	Continuous index: average score of seven binary items about the acceptability of risky sexual behavior (multiplied by 100 to represent the percentage of items agreed with). Average scores range from 0 to 100 with higher values representing more support for risky behavior.
Domain: Motivation	
Motivation to delay childbearing	Continuous index: average of three items about motivation to delay childbearing. Average scores range from 1 to 4 with higher values representing greater levels of motivation.
Domain: Intentions (in next 12 months)	
Intention to have oral sex in the next year	Single item scored 0 or 1, with 1 representing stronger intention.
Intention to have sexual intercourse in the next year	Single item scored 0 or 1, with 1 representing stronger intention.
Intention to use a condom if having sexual intercourse in the next year	Single item scored 0 or 1, with 1 representing stronger intention.
Intention to use birth control if having sexual intercourse in the next year	Single item scored 0 or 1, with 1 representing stronger intention.

Measure	Definition
Domain: Skills	
Condom negotiation skills	Continuous index: average of responses to seven questions about perceived ability to obtain and negotiate the use of condoms. Average scores range from 1 to 4 with higher values representing greater certainty about condom negotiation skills.
Refusal skills	Continuous index: average of responses to six questions about perceived ability to refuse to engage in risky sexual behavior. Average scores range from 1 to 4 with higher values representing greater certainty about refusal skills.
SEXUAL ACTIVITY AND RISK BEHAVIOR	
Currently sexually active (in last 90 days)[a] Sexual intercourse in the last 90 days Oral sex in the last 90 days	Single items, yes (1) / no (0) answer.
Initiation of sexual activity	For those who were not sexually active at baseline, indicates whether they became sexually active between baseline and follow-up. Single item, yes (1) / no (0) answer.
Sexual intercourse without birth control (in last 90 days)[a] Sexual intercourse without a condom (in last 90 days) Oral sex without a condom (in last 90 days)	Single items, yes (1) / no (0) answer.

[a] Designated as a confirmatory outcome, as discussed below in Section 3.4.5.

3.4 Analytic Approach

The impact analysis examines the extent to which *RtR* affected each of the study's outcomes. In testing for these effects, we use two-tailed hypothesis test procedures, because we do not want to rule out the possibility that the intervention might adversely affect one or more of the outcomes.

Our basic strategy for estimating program impacts is to compare the outcomes of treatment and control group members using a regression framework, in which we include baseline covariates to increase statistical precision (i.e., reduce the standard errors) of the impact estimates for a given sample size (Orr, 1999) and reduce attrition bias from missing data (see Puma et al., 2009).

3.4.1 Estimation of Impacts for the Full Sample

We report impact estimates that are pooled across the three *RtR* replication sites. OAH's requirements to define, measure, and adhere with fidelity to the program model mean that each of the three replication sites implemented the same core program elements. The random assignment and data collection procedures were also the same across all sites. These design elements ensure that impact estimates pooled at the program level represent rigorous tests of a well-defined and consistently implemented program model.

For this evaluation of *RtR*, classrooms were randomly assigned within random assignment blocks to treatment or control conditions. The random assignment blocks in each site comprised groups of classes within schools and semesters that were similar to one another in the ages and grades of students in the classes. In order to ensure the impacts were estimated by comparing treatment boys to control boys and treatment girls to control girls within sites, schools and semesters, block dummies were included in the analytic model that further split the original randomization blocks by gender to form

site/school/semester/gender blocks. This forces the analytic model to estimate an overall average impact that is a precision weighted average of estimated treatment effects within the within site, school, semester, and gender blocks. For each outcome, we estimate a regression model that accounts for the clustering of students within classrooms, which increases the standard errors of the impact estimates. To account for this form of clustering, we use a multi-level modeling approach that has the basic structure of equations 1-3 below.

In this model, individual outcomes are modeled at Level 1, while Level 2 represents the unit of random assignment (or "cluster"). The Level 1 model includes individual-level demographics and baseline measures as covariates, and dummies for the site/school/semester/gender blocks. The block dummies appear in the level-1 model because there are male and female blocks within classes. The Level 2 model includes a treatment indicator and random intercept terms for classes to account for correlation of individuals within classes. Information about sites is contained within the block dummies. There are no specific model terms for sites because the block dummies are linear combinations of the site indicators.

(1) Level 1: $$Y_{ij} = \beta_{0j} + \sum_{k=1}^{K} \beta_{kj} X_{kij} + \sum_{m=1}^{M} \delta_{mj} D_{mij} + \varepsilon_{ij}$$

(2) Level 2: $$\beta_{0j} = \gamma_{00} + \gamma_{01} T_j + \mu_{0j}$$

(3) Level 2: $$\beta_{kj} = \gamma_{k0}$$

(4) Level 2: $$\delta_{mj} = \gamma_{(k+m)0}$$

In this model:[12]

Y_{ij} is the outcome of interest (e.g., sexual intercourse without a condom) for the i^{th} student in the j^{th} class, m^{th} site/school/semester/gender block;

T_j is a dummy variable equal to 1 if class j was assigned to the treatment group and 0 otherwise;

X_{kij} is the k^{th} baseline characteristic or covariate for individual i. These include baseline age, grade, race/ethnicity (Black, White, Hispanic (omitted), other), risk behaviors (smoking, alcohol use, marijuana use), baseline sexual activity (ever sexually active), baseline knowledge (pregnancy and STI risk), baseline intentions (intention to have oral sex and sexual intercourse), and the baseline measure of the outcome when available.

D_{mij} is a dummy variable equal to 1 if student i was within the m^{th} site/school/semester/gender block and 0 otherwise.

[12] The analyses presented in this report used multi-level linear models. A set of robustness analyses were conducted using multilevel logistic regression models and using multi-level linear models with heteroskedasticity robust standard errors for binary outcomes (Constantine et al, 2009, Gleason et al., 2010). There were no substantive differences in the inferences that results from any of the three modeling approaches.

The coefficient γ_{01} is interpreted as the average pooled impact of the program on the outcome. Additionally, β and γ are coefficients to be estimated, and ε_{ijs} and μ_{0js} are random terms. The regression covariates, X_{kij}, reflect the influence of background characteristics on the control group mean. It is important to note that this model specification treats site/school/semester/gender blocks and the treatment effects as fixed as opposed to random, which is consistent with how the replication sites are chosen and how the results of the study will be interpreted.[13]

Equations 1-4 estimate the impact of access to *RtR*. Because of the random assignment design, the crucial difference between the treatment and control groups is ***access*** to *RtR* services: Individuals in the treatment group had access to program services and possibly other, potentially similar, services available in the school or community (e.g., clinics), while control group members had access to only those other services in the school or community. In the evaluation literature, the estimate of the average impact of access is referred to as the intent-to-treat (ITT) impact parameter. It measures the impact of having the opportunity to participate in the intervention on the outcomes under consideration for the average individual given access, not the average impact on program group members who actually participate in the intervention.

Finally, we report impact findings in tables showing the regression-adjusted treatment group mean, the unadjusted control group mean, and the regression-adjusted impact (and p-value). For binary outcomes (e.g., condom use) and outcomes measured on a 0-100 percent scale, we report impacts as percentage point differences between the treatment and control group means. For all other outcomes, we show impact estimates in their original metric and additionally convert impact estimates to standardized effect sizes (SES) by dividing the impact estimate by the pooled standard deviation of the treatment and control groups, and we report these in a separate column.

3.4.2 Site-Level Analyses

In addition to estimating impacts pooled across the three replication sites, we estimate impacts for each site separately and test for differences in impact across the three sites. We implement these analyses by including treatment by site interaction terms in the model (i.e., Equation 1) and testing for the joint significance of the interaction terms.[14] When statistically significant differences are found between sites for one or more outcomes, we discuss these differences.[15] Site-specific impact estimates for all outcomes are presented in Appendix A.

3.4.3 Subgroup Analyses

In addition to the overall pooled impacts and site-level impacts, we estimate impacts for key subgroups of participants (based on age, gender, race/ethnicity, and sexual experience at baseline) and test for

[13] Because replication sites were selected as a purposive sample, not randomly selected from a larger population of sites, we do not consider a random treatment effects model to be appropriate for drawing inferences in this sample (Schochet, 2008a, p. 70).

[14] For the treatment-by-site interaction, a two degrees-of-freedom F test was used.

[15] The purpose of testing for differences across sites before discussing results in the main text is to guard against over-interpretation of spurious findings, some of which would be expected by chance in such a large group of outcomes. The basic idea behind the strategy of discussing site-specific impacts only when differences are found is that it is only credible to report an impact in one site – but not in another – if there is a significant difference between the two sites. The site-specific results in Appendix A are not adjusted for multiple comparisons and any significant findings reported there should therefore be interpreted with caution.

analyses by including subgroup indicators and treatment by subgroup interaction terms in Level 1 of the model (i.e., Equation 1) and testing for significance of the interaction term.[16]

To guard against potential over-interpretation of results among the very large number of subgroup estimates, we present impact estimates for individual subgroups only when there is a statistically significant difference between subgroups; for example, the impact would be presented for the subgroup of boys only if there were a statistically significant difference in impacts between boys and girls (see Appendix B).

3.4.4 Handling Missing Data

We used monetary incentives (gift cards) and intensive tracking to achieve the maximum possible response rate for the short-term follow-up survey for both treatment and control groups, and have achieved very high response rates in each of the replication sites (see Exhibit 3.3).

We use case deletion for the few instances of missing outcome data (Puma et al., 2009). Dummy-variable adjustment is used in regression models to account for missing covariates. In the dummy variable adjustment method, missing covariate values are set to a constant and indicators (or dummy variables) for such values are added to the impact analysis model (Puma et al., 2009).

3.4.5 Addressing Multiple Comparisons

Ongoing developments in the statistical analysis of the results of randomized trials emphasize that conventional statistical tests and confidence intervals apply to a single outcome. When analysts look over multiple outcomes for any statistically significant finding, the appropriate critical t-values are much higher; that is, effects that appear to be statistically different from zero are not truly different from zero. In the literature, this is known as the problem of "multiple comparisons." Current guidance on how to approach this multiple comparison problem recommends distinguishing two categories of analyses (Schochet, 2008b). One—called "confirmatory tests"—includes a small number of critical outcome domains for which it is important to adjust error probabilities for multiplicity. Confirmatory analysis uses a high standard of evidence for deciding whether an intervention has had its intended effect, in order for its findings to be considered conclusive rather than merely suggestive. A second category includes "exploratory tests" for which there is generally higher tolerance of errors and for which multiplicity adjustments may or may not be made.

For this report, the impact analysis team pre-specified a multiple comparisons strategy that spans the short- and long-term impact reports and includes confirmatory and exploratory analyses. The confirmatory analysis seeks convincing evidence that *RtR* improved participants' behavioral outcomes past the end of the program. Before analyzing data, the team pre-specified a small number of outcomes in three "domains," or sets of similar constructs, as part of the overall analytic strategy for both reports. The three confirmatory outcome domains are: *youth sexual behavior at the short-term follow-up, recent sexual behavior at the longer-term follow-up*, and *pregnancy*.

To control for multiple comparisons within each of the confirmatory domains, we apply a formal multiple comparisons correction (in particular, a Benjamini-Hochberg correction as described in Appendix G of the What Works Clearinghouse Procedures and Standards Handbook, version 3.0 (U.S. Department of Education, 2014), which controls for the false positive rate by adjusting *P*-value thresholds). The

[16] For the treatment*race/ethnicity interaction, a 3 degrees of freedom F test was used.

correction does not affect the p-values that appear in tables of results, but it does change the interpretation of statistical significance. In particular, it raises the bar for rejecting the null hypothesis.

Two outcomes in this short-term report, **currently sexually active** (engaged in sexual intercourse, oral sex, and/or anal sex in the last 90 days) and **sexual intercourse without birth control** (engaged in sexual intercourse without a condom or other birth control in the last 90 days), were pre-specified as key outcomes in one of the study's three confirmatory outcome domains, *youth sexual behavior at the short-term follow-up.* The other two domains, *recent sexual behavior at the longer-term follow-up* and *pregnancy,* will be analyzed in the final, longer-term impact report, along with the findings presented here.

The exploratory analysis encompasses all other outcomes and research interests in the short-term report, for example, impacts on intermediate outcomes and impacts on other behavioral outcomes. Given the large number of hypothesis tests that constitute the exploratory analysis, some false positive findings are to be expected. We do not make formal adjustments for multiple comparisons when reporting on statistical significance. However, to aid in interpretation, we specify the number of tests that were conducted (within domains) and the number of false rejections that would be expected given the number of tests if there were no impact of treatment.

4. Results

This study is designed to determine whether *RtR* helps young people develop the knowledge, attitudes, and skills to act in ways that ultimately protect them from the potential consequences of sexual risk behavior, such as STIs and unintended pregnancy. The program, when delivered with fidelity, is intended to provide information and affect potential intermediate outcomes such as knowledge and understanding of reproductive health and avoidance of sexual risk; attitudes toward using protection; motivation to delay pregnancy; intentions to become sexually active and use protection; and skills needed to avoid sexual risk. The ultimate goals are reduced rates of unprotected sexual activity and unplanned pregnancy.

The short-term findings (12 months post-baseline) discussed here suggest that:

- The *RtR* program was implemented as intended.
- It was effective in increasing knowledge and attitudes (toward using protection).
- However, these changes did not lead to overall improvements in reported sexual risk behaviors.

In this section, we expand on our conclusion that the program was indeed implemented with fidelity across replication sites, and then discuss findings for the full study sample and for individual sites, as well as any important findings for specific subgroups of youth (i.e., sexual experience at baseline, age, gender, and race/ethnicity).

In addition to the exhibits in this section, tables documenting the site-level analyses can be found in Appendix A, and the corresponding tables documenting subgroup analyses in Appendix B.

4.1 Program Implementation

As noted in Section 1.3, a separate report will provide a detailed account of the implementation of *RtR* in the three replication sites. That implementation report serves two important purposes: (1) to help explain the findings of the Impact Study and (2) to offer lessons learned to help those planning to use the *RtR* program in the future.

What we have learned from the Implementation Study that is directly relevant for this short-term impact report is that the program was well implemented across the three replications. The three grantees hired staff with appropriate background experience and skills to deliver the program; all received training approved by the developer; the program was implemented with fidelity to its core elements and without modifications that threatened those core elements; and attendance was generally strong.

4.1.1 Staff Hiring and Training

The three grantees were consistent in the types of experience and skills they sought when hiring health educators (or identifying one or more from current agency staff). Experience working with adolescents and in sexual health, and comfort in addressing adolescent sexual health issues, were considered essential. All of the staff received the official training provided by the curriculum distributor and approved by the developer. Grantees offered additional training for staff and encouraged them to attend training sessions offered by OAH, as well as state or local agencies and institutions. Staff retention was high.

4.1.2 Implementing the Program with Fidelity

As part of the TPP Program, OAH stipulated that grantees maintain fidelity to the core components of their chosen program model, and provided grantees guidance on making minor adaptations (all of which had to be approved by OAH before they could be implemented). There was an accompanying requirement that grantees develop a plan to monitor fidelity of implementation and continued adherence to the core program model.

For *RtR*, fidelity monitoring log templates were provided by the developer to help the grantees collect this information. Health educators were required to complete a fidelity log for each module delivered. In addition, OAH provided observation protocols, to be used by supervisory staff on a regular schedule, that allowed an assessment of the **quality** of the sessions. Data from the logs and observations were aggregated and used by program supervisory staff to identify areas where improvement was needed. Aggregate data were delivered to OAH every six months and summarized to provide a basis for subsequent discussions between OAH program officers and the grantees. All of these activities were intended to guide implementation and ensure not just fidelity but a degree of uniformity across sites replicating the same program model.

The approved modifications described in Section 2.1 were not viewed by OAH or the program developer as affecting implementation of the core elements of the program model. Each of the replication sites successfully delivered the intervention to students with fidelity. Nevertheless, grantees discovered they needed to develop strategies to address implementation challenges created by factors external to the program. SDYS, faced with very large class sizes in some schools as a result of budget cuts, responded by assigning two health educators to each class. Because space in those schools is at a premium, it was impossible to break a large class into two groups, a preferable solution. Space issues also dogged LifeWorks health educators; often, they were not assigned space until the day of the class, which sometimes reduced the time available for the session. BFL enjoyed excellent relationships with staff in all six schools, but had to deal with student absences in some higher-risk schools.

4.1.3 Participant Attendance and Engagement

Grantees were required to collect and report students' attendance (by session) using attendance logs. SDYS reported that 85 percent of students attended at least 75 percent of classes; LifeWorks reported 47 percent of students; and BFL reported 73 percent.

Abt's independent observations and focus group sessions with students suggest that students actively participated in program sessions and acquired new information from the program.

4.2 Characteristics of the Student Sample at Baseline

4.2.1 Study Sample at Baseline

Baseline characteristics of the overall *RtR* study sample and for each replication site are presented in Exhibit 4.1. At baseline, students in the study sample were, on average, 14½ years old. Students in the SDYS sample were, on average, a year or more younger than students in the other two sites.

Overall, more than a third of study participants were non-Hispanic, Black; more than 40 percent were Hispanic, and the remainder were non-Hispanic White or Other. However, there were large differences among the three replication sites in the racial/ethnic composition of the study sample. The BFL sample was almost entirely Black; whereas Black students comprised less than 10 percent of the sample in the

two other sites. By contrast, Hispanic students comprised about two-thirds of the sample in SDYS and LifeWorks, compared with fewer than 3 percent in BFL.

Across all three replication sites, more than 90 percent of students lived with at least one biological parent. Almost two-thirds reported feeling close to and cared for by their mothers, and almost half reported feeling the same about their fathers.

Almost half of the sample reported ever drinking alcohol; less than one-third reported ever using marijuana; and one-fifth reported ever smoking cigarettes. In San Diego, where the sample was younger, significantly fewer students had engaged in any of these risk behaviors.

Just over half of the students (51 percent) in the overall sample demonstrated an accurate understanding of pregnancy risk, while a somewhat smaller proportion (44 percent) understood STI risks. On both topics, the LifeWorks sample was better informed than students in the other two sites. Across all three sites, students reported relatively supportive attitudes toward using protection. Overall, almost one-third of the sample expressed an intention to engage in oral sex in the next 12 months, and 40 percent intended to have sexual intercourse in the same period. Significantly fewer in the SDYS sample expressed these intentions compared with students in the other two sites. Across the sample, students reported strong intentions to use protection if they were to have sexual intercourse in the next 12 months.

There were substantial and significant differences across the three sites in the extent to which students had engaged in sexual activity and sexual risk behavior before they entered the study. While less than a third of the overall sample had ever been sexually active, and one-fifth were currently sexually active, the SDYS sample was strikingly less sexually experienced: just 12 percent had ever been sexually active and less than 10 percent were currently sexually active. Students in BFL consistently reported the highest levels of sexual activity. This same pattern repeated for sexual risk behaviors. The proportions of students who had engaged in unprotected sex were consistently lowest in the SDYS sample and highest in the BFL sample.

Almost 60 percent of the sample reported receiving information on birth control methods in the year prior to the study, but less than half reported receiving information on where to obtain it. More than 80 percent of students reported receiving some information on STIs, probably from one or more health classes delivering mandated lessons on HIV. Almost all (88 percent) had been exposed to information about reproductive anatomy ("how babies are made"), and more than 70 percent had received information about abstinence. There was some, but not consistent, variability across sites in the proportions of students who had been exposed to the kind of information offered by *RtR* before entering the study.

Exhibit 4.1: Baseline Characteristics by Site

Measure	Better Family Life	LifeWorks	San Diego Youth Services	*RtR* Overall	p-value for the Test of Differences across Sites [a]
Demographic characteristics					
Age (years)					
Mean	14.74	15.14	13.75	14.54	0.000 ***
Grade	9.19	9.80	8.82	9.26	0.000 ***
Gender (percent female)	47.61	48.65	52.63	49.61	0.079

Measure	Better Family Life	LifeWorks	San Diego Youth Services	*RtR* Overall	p-value for the Test of Differences across Sites [a]
Race/ethnicity [b] (percent)					
Hispanic	2.55	63.42	68.38	43.77	0.000 ***
Black	90.01	8.68	5.03	35.92	0.000 ***
White	0.85	22.16	10.17	10.71	0.000 ***
Other	6.59	5.74	16.42	9.59	0.000 ***
Family structure and relationships					
Lives with biological parent/s	91.68	93.33	92.80	92.58	0.400
Feels very close to and cared for by father	45.90	44.20	48.39	46.16	0.251
Feels very close to and cared for by mother	67.73	61.63	64.34	64.66	0.028 *
Risk behaviors					
Ever smoked cigarettes	19.50	29.47	13.90	20.87	0.000 ***
Ever drank alcohol	48.92	55.38	33.29	45.86	0.00 0 ***
Ever used marijuana	34.70	39.57	17.96	30.78	0.000 ***
Knowledge, attitudes and intentions					
Knowledge of pregnancy risk	45.65	58.70	49.60	51.11	0.000 ***
Knowledge of STI transmission	44.18	48.96	38.62	43.87	0.000 ***
Attitudes toward protection (1 = least supportive, 4 = most supportive)	3.03	3.03	3.06	3.04	0.221
Intentions to have oral sex in the next 12 months	34.17	37.47	20.18	30.58	0.000 ***
Intentions to have sexual intercourse in the next 12 months	52.19	45.97	24.08	40.87	0.000 ***
Intentions to use a condom if they were to have sexual intercourse in the next 12 months	93.85	94.14	95.40	94.44	0.328
Intentions to use birth control if they were to have sexual intercourse in the next 12 months	87.86	88.16	93.68	89.84	0.000 ***
Sexual behavior					
Ever sexually active [c]	47.37	36.17	11.74	31.97	0.000 ***
Currently sexually active (in the last 90 days) [b]	31.50	22.16	7.31	20.51	0.000 ***
Sexual intercourse in the last 90 days	28.07	19.76	6.16	18.17	0.000 ***
Oral sex in the last 90 days	19.10	16.09	5.70	13.71	0.000 ***
Sexual risk					
Sexual intercourse without a condom in the last 90 days	12.83	12.34	2.44	9.24	0.000 ***
Oral sex without a condom in the last 90 days	14.60	14.29	5.12	11.37	0.000 ***
Sexual intercourse without birth control in the last 90 days	8.88	7.78	1.86	6.21	0.000 ***
Baseline exposure to sexual health information [d]					
Relationships or marriage	79.29	81.54	81.33	80.68	0.409
Abstinence from sex	74.79	70.07	71.49	72.20	0.074
Birth control methods	55.53	59.36	60.97	58.54	0.055
Where to obtain birth control	46.18	45.66	46.94	46.26	0.868
Sexually transmitted infections	87.77	82.49	82.00	84.18	0.001 ***

Measure	Better Family Life	LifeWorks	San Diego Youth Services	*RtR* Overall	p-value for the Test of Differences across Sites [a]
How to talk with partner about sex and birth control	58.82	48.93	47.45	51.93	0.000 ***
How to say no to sex	76.67	68.13	71.23	72.15	0.000 ***
How babies are made	87.18	86.15	90.24	87.86	0.025 *

Source: Baseline survey administered prior to randomization.

Notes: Results in this table are based on 2,604–2,689 respondents (for *RtR* overall) who provided valid survey responses to relevant items, except for the items measuring how close students feel to their mothers (n=2,592) and fathers (n=2,368) and intentions to use birth control (n=2,580). The items that compose measures of attitudes toward risky behavior, motivation to delay childbearing, refusal skills, and condom negotiation skills were not asked at baseline.

[a] Test results from an analysis of variance testing the null hypothesis that the means of the variable indicated in the row are equivalent among the three sites.

[b] Racial/ethnic categories are Hispanic, Black non-Hispanic, White non-Hispanic, and Other non-Hispanic, where Other is defined as Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, multiracial, or undisclosed race.

[c] Sexual activity is defined differently across grantees. In Better Family Life, sexual activity refers to sexual intercourse, oral sex, and anal sex. In LifeWorks and San Diego Youth Services, students were not asked about anal sex.

[d] Questions refer to information students received in the 12 months prior to the survey administration.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

4.2.2 Comparability of the Groups at Baseline

Although the characteristics of study participants differed significantly across the three replication sites (reflecting the differences in student populations in those sites), there were no significant differences between students assigned to the treatment group and students assigned to the control group.

Baseline treatment-control differences were estimated using a series of models with the same structural components as the impact model in Equation 1 (i.e., the same block indicators and treatment group indicator), but where in each model one baseline characteristic (from among those in Exhibit 4.1) served as the dependent variable, and where the other covariates used in the impact model were omitted. In this approach, the coefficient for the treatment indicator is the treatment-control difference on the pre-test measure. None of these differences was significant. (See Appendix Table D.1.)

4.3 Program Impacts on Exposure to Sexual Health Information

In each of the replication sites, the *RtR* curriculum represented a way to provide youth with sexual health information that may or may not be available from other sources. Each of the replications sought to supplement what students typically get in schools (often a maximum of two sessions in a high school health class) and address a perceived lack of services and information.

We anticipated that after participation in *RtR*, students in the treatment group would be more likely to report exposure to sexual health–related information than would their counterparts in the control group. As expected, *RtR* had a statistically significant and positive effect on students' exposure to information about sexual health topics. That is, 12 months after the *RtR* program began:

There were moderate to large significant and positive impacts on five out of eight measures of students' exposure to information about abstinence, contraception methods, where to obtain contraception, how to talk to a partner about sex and contraception, and how to say no to sex (Exhibit 4.2).[17]

There were no significant differences in impacts between sites (see Appendix Table A.1).

Exhibit 4.2: Short-term Impacts on Exposure to Sexual Health Information

Outcome	Adjusted Treatment Mean [a]	Unadjusted Control Mean	Treatment Effect [b]	p-value
Percentage of respondents that reported receiving information on the following topics:				
Relationships or marriage	80.71	80.34	0.36	0.837
Abstinence from sex	74.51	67.15	7.36 ***	0.000
Birth control methods	72.56	63.07	9.49 ***	0.000
Where to obtain birth control	69.95	57.87	12.08 ***	0.000
Sexually transmitted infections	84.11	81.72	2.39	0.136
How to talk with partner about sex and birth control	76.71	65.52	11.19 ***	0.000
How to say no to sex	82.23	72.08	10.15 ***	0.000
How babies are made	87.94	86.61	1.33	0.305

Source: Follow-up survey administered 12 months after baseline.

Notes: Questions refer to information received in the 12 months prior to the survey administration. Results in this table are based on 2,682 - 2,688 respondents who provided valid survey responses to relevant items.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

** p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).*

Subgroup Differences. There were differences in impacts on exposure to information by specific subgroups defined by: sexual experience at baseline, student age, gender, and race/ethnicity (see Appendix Table B.1).

> ***Sexual experience at baseline.*** Effects on exposure to information about abstinence, contraception methods, where to obtain contraception, and communicating with a partner were almost entirely located in the large group of students who had no sexual experience at baseline. There was approximately a 10 to 15 percentage point difference between treatment and control group students who had never been sexually active at baseline in terms of whether or not they had been exposed to this information.
>
> ***Age.*** Regarding information about where to obtain contraception, although the program had a significant effect on students across the age range, the impact on younger students (15.2 percentage point difference) was twice as large as the impact on older students (7.9 percentage point difference).

[17] In the absence of a true program impact, with eight tests and a significance criterion of p<0.05, the expected number of findings that would be significant by chance alone is less than one.

Gender. Although both male and female *RtR* participants were significantly more likely to report exposure to information about how to say no to sex than their control group counterparts, the impact was more than three times as large for males (15.3 percentage point difference) as for females (4.8 percentage point difference).

Race/ethnicity. No overall effect on knowledge of reproductive anatomy ("how babies are made") was found (probably because the vast majority of students in both groups were exposed to this information), but *RtR* did have a significant impact on students who categorized themselves as "Other" race. There was a 9 percentage point difference between treatment and control group students of "Other" race in terms of whether or not they had been exposed to this information.

4.4 Intermediate Outcomes

The *RtR* program's theory of change (see logic model in Exhibit 2.2) specifies intermediate outcomes that the model predicts will influence behavior—namely, **knowledge** and understanding of sexual risk behaviors and their prevention or avoidance; **attitudes** toward protection and risk behaviors; **motivation** and **intentions** to engage in sexual activity; and refusal and negotiation **skills**.

- We find evidence that *RtR* had positive impacts on knowledge and attitudes.
- There were no overall program effects on motivation, intentions, or skills.

4.4.1 Knowledge

RtR had statistically significant impacts on knowledge of pregnancy risk and knowledge of STI risk.

Knowledge of Pregnancy Risk. The effect of *RtR* on knowledge of pregnancy risk was assessed using a composite measure that combines four survey items on topics such as the effectiveness of condoms and birth control in preventing pregnancy. Compared with control group students, treatment group students scored significantly higher on the composite measure and were significantly more likely to give the correct answers for three of the four individual items.[18] Exhibit 4.3 shows the findings for the composite measure and the individual items.

Knowledge of STI Risk. *RtR* also had statistically significant positive impacts on a composite measure and on individual survey items measuring student knowledge of STI risk. Students in the treatment group were significantly more likely to correctly answer questions about the effectiveness of birth control and condoms in preventing HIV and other STIs. There were also impacts on some individual items assessing students' understanding of the transmission of STIs and the consequences of sexual activity.

Of the items in the composite measure, those on which the program had no impact were: (1) understanding that some STIs place you at greater risk for HIV and (2) understanding that a person who looks and feels healthy can transmit an STI. In the case of the latter item, more than 75 percent of students in both groups knew the correct answer. It is also worth noting that, although the program had a significant impact on students' understanding of the effectiveness of condoms (and the ineffectiveness of birth control pills) in preventing both gonorrhea and HIV, even in the treatment group, the percentages of

[18] There were significant, positive effects of *RtR* on four out of five measures of knowledge of pregnancy risk (Exhibit 4.3). In the absence of a true program impact, with five tests and a significance criterion of $p<0.05$ the expected number of findings that would be significant by chance alone is less than one.

correct answers were quite low—one half or fewer of the students in either group.[19] Exhibit 4.3 shows the findings for the composite measure and the individual survey items.

There were no significant differences in short-term impacts on knowledge of pregnancy risk or knowledge of STI risk between sites (see Appendix Table A.2) or by subgroup.

Exhibit 4.3: Short-term Impacts on Knowledge of Pregnancy Risk and STI Risk

Outcome	Adjusted Treatment Mean [a]	Unadjusted Control Mean	Treatment Effect [b]	p-value
Knowledge of pregnancy risk (percent of items respondent answered correctly) [c]	65.55	61.55	4.01***	0.000
Percent of respondents correctly answering each item				
Used correctly, how much can birth control pills reduce pregnancy risk?	59.03	54.85	4.18*	0.030
Used correctly, how much can condoms reduce pregnancy risk?	63.43	59.74	3.68*	0.045
A couple that has had unprotected sex and not gotten pregnant does not have to worry about getting pregnant.	80.59	80.00	0.59	0.697
A woman is protected from pregnancy the day she begins taking the pill.	59.37	51.59	7.78***	0.000
Knowledge of STI risk (percent of items respondent answered correctly) [d]	60.47	56.21	4.26***	0.000
Percent of respondents correctly answering each item				
Once you are infected with HIV you are infected for life	76.93	72.53	4.39*	0.011
There is a vaccine to prevent girls from getting HPV	45.46	41.63	3.82*	0.050
All STDs/STIs can be cured by taking medicine	66.68	62.06	4.62*	0.011
A person with an STD/STI who looks and feels healthy cannot transmit the infection to others	79.33	76.65	2.68	0.088
Some STDs/STIs put you at greater risk of HIV	61.27	62.06	-0.80	0.718
About 1 out of 4 sexually active teens gets an STD/STI every year	63.40	59.31	4.09*	0.034
You can get an STD/STI from having oral sex	71.74	66.61	5.13**	0.008
Used correctly, how much can condoms decrease the risk of HIV?	51.83	47.21	4.62*	0.016
You can't get infected with HIV even if you have sex only once or twice w/o a condom.	64.94	61.20	3.74*	0.045
Used correctly, how much can condoms decrease the risk of gonorrhea?	40.54	33.05	7.49***	0.000
Used correctly, how much can birth control pills decrease the risk of HIV?	52.16	48.24	3.92*	0.030
Used correctly, how much can birth control pills decrease the risk of gonorrhea	51.25	43.95	7.30***	0.000

[19] There were significant, positive effects on 11 out of 13 measures of knowledge of STI risk (Exhibit 4.3). In the absence of a true program impact, with 13 tests and a significance criterion of $p<0.05$ the expected number of findings that would be significant by chance alone is about one.

Source: Follow-up survey administered 12 months after baseline.
Notes: Results in this table are based on 2,689 respondents who provided valid survey responses to relevant items.
[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).
[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.
[c] Score based on the four items below. Values shown represent the average percent of items answered correctly by respondent for each group. Alpha coefficient = 0.52.
[d] Score based on the 12 items below. Values shown represent the average percent of items answered correctly by respondent for each group. Alpha coefficient = 0.68.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

4.4.2 Attitudes

RtR had statistically significant impacts on students' attitudes toward using birth control or condoms.

Attitudes Toward Protection. For the composite measure on attitudes toward protection (comprising both birth control and condoms), students in the treatment group had significantly more positive (and protective) attitudes. On a scale ranging from 1 to 4, with higher values indicating more positive attitudes, the mean for the treatment group was 3.18 while the mean for the control group was 3.13. Of the 12 items that comprise this composite measure, there were significant impacts on students' belief that both birth control and condoms are "pretty easy to get." In addition, the program was effective in dispelling negative attitudes regarding potential side effects of birth control pills and about the effect of condoms on sexual pleasure. *RtR* also had a significant impact in the intended direction on students' attitudes about the role of birth control in making sex safer (but not on the role of condoms in making sex safer, probably because students in both groups scored high on the item) and on whether using condoms means you don't trust your partner (Exhibit 4.4).[20]

Attitudes Toward Risky Sexual Behavior. *RtR* had no statistically significant impacts on the composite measure of attitudes toward risky sexual behavior or on any of seven individual items measuring such attitudes (Exhibit 4.5). At 12 months after baseline, the overwhelming majority of students in both the treatment and control groups rejected the view that risky behavior is acceptable. On average, in both groups, only about 5 percent expressed support for risky behavior.

Site-Level Differences. Site-level analyses suggest that the SDYS replication site was largely responsible for the overall impact on attitudes toward protection observed in the full sample. The positive program impact (SES = .30) on the composite measure of attitudes toward protection was highly significant in SDYS, while there were no significant impacts on the measure in the other two sites (Appendix Table A.3).There were no site-level differences in the impacts of *RtR* on attitudes toward risky sexual behavior (Appendix Table A.4).

Subgroup Differences. There were no subgroup differences in the impacts of *RtR* on attitudes toward protection. However, subgroup analyses showed that there were significant differences in the impact of *RtR* on attitudes towards risky sexual behavior by age and race/ethnicity (Appendix Table B.2).

[20] There were significant, positive effects on 7 out of 13 measures of attitudes toward protection (Exhibit 4.4). In the absence of a true program impact, with 13 tests and a significance criterion of p<0.05 the expected number of findings that would be significant by chance alone is about one.

Age. There was a significant adverse program impact on attitudes toward risky sexual behavior for students aged 15 and older (2.3 percentage point difference). That is, older students who received the program had more positive attitudes toward risky sexual behavior compared with older students in the control group.

Race/ethnicity. There was a significant adverse program impact on attitudes toward risky sexual behavior for White students. That is, White program participants had more positive attitudes toward risky sexual behavior compared with White students in the control group (5.7 percentage point difference).

Exhibit 4.4: Short-term Impacts on Attitudes toward Protection

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	p-value	Effect Size[c]
Attitudes toward protection[d]	3.18	3.13	0.05***	0.000	0.13
Birth control pills should always be used if a person your age has sexual intercourse	3.23	3.22	0.01	0.674	0.02
Birth control is too much trouble to use (reverse-coded)	3.08	3.09	-0.02	0.608	-0.02
Birth control is pretty easy to get	2.87	2.74	0.12***	0.000	0.15
Birth control is important to make sex safer	3.26	3.18	0.08*	0.011	0.10
Birth control has too many side effects (reverse-coded)	2.60	2.53	0.07*	0.037	0.09
Using birth control is morally wrong (reverse-coded)	3.22	3.18	0.04	0.207	0.05
Condoms are too much trouble to use (reversed-coded)	3.31	3.27	0.04	0.169	0.05
Condoms are pretty easy to get	3.33	3.24	0.09**	0.005	0.11
Condoms are important to make sex safer	3.62	3.62	0.01	0.808	0.01
Using condoms means you don't trust your partner (reverse-coded)	3.38	3.30	0.08*	0.011	0.10
Using condoms is morally wrong (reverse-coded)	3.52	3.47	0.05	0.06	0.08
Condoms decrease sexual pleasure (reverse-coded)	2.69	2.62	0.07*	0.023	0.09

Source: Follow-up survey administered 12 months after baseline.

Notes: Results in this table are based on 2,598 - 2,688 respondents who provided valid survey responses to relevant items.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed in the original metric of the outcome variable. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[c] The effect size is the standardized effect size of the difference, which is the "treatment effect" divided by the pooled standard deviation of the treatment and control groups.

[d] This construct averages responses to 12 items (shown in table) on attitudes towards condoms and birth control. Possible values range from 1 to 4 with higher values indicating more positive attitudes toward protection (alpha coefficient = 0.75).

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

Exhibit 4.5: Short-Term Impacts on Attitudes toward Risky Sexual Behavior

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	p-value
Attitudes toward risky sexual behavior [c]	5.32	4.53	0.80	0.161
Percent of respondents who agreed with each item				
It's OK to have sex with someone on your first date.	7.19	6.56	0.63	0.528
It's OK to have sex with someone the same night you meet them.	6.16	5.61	0.55	0.545
It's OK to have sex with several different people in the same month.	4.77	5.01	-0.24	0.785
It's okay to have sex without protection.	4.48	3.20	1.28	0.090
It's OK to have sex with someone when you know they are someone else's girlfriend/boyfriend.	4.69	3.28	1.41	0.082
It's OK to have sex with someone if you are drunk or high.	5.91	5.09	0.82	0.362
It's OK to have sex with someone if you know they are drunk or high.	4.16	2.94	1.23	0.091

Source: Follow-up survey administered 12 months after baseline.

Notes: Results in this table are based on 2,675 respondents who provided valid survey responses to relevant items.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression-adjusted impact estimate (treatment effect).

[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[c] Score based on the seven items (shown below) represents the average percent of items agreed with by respondent for each group (alpha coefficient = 0. 82).

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

4.4.3 Motivation

Students in both the treatment and control groups were highly motivated to delay childbearing. There were no differences between the two groups on the composite measure or on any of three individual items. Participants in both groups indicated a belief in the importance of delaying childbearing until personal goals have been achieved (Exhibit 4.6).

Site-Level Differences. Site-level analyses revealed a small (SES = −.18) but significant unintended effect of the program in the BFL sample (Appendix Table A.5). In BFL, students who had participated in *RtR* were less motivated to delay childbearing than were those in the control group.

There were no subgroup differences in the short-term impacts of *RtR* on motivation to delay childbearing.

Exhibit 4.6: Short-term Impacts on Motivation to Delay Childbearing

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	p-value	Effect Size[c]
Motivation to delay childbearing[d]	3.68	3.68	-0.01	0.741	-0.01
You have goals you want to accomplish before having a child.	3.62	3.65	-0.02	0.448	-0.03
It is important for you to finish school before you have a child.	3.69	3.70	0.00	0.884	-0.01
It is important to have a job and a stable income before you have a child.	3.71	3.71	0.00	0.983	0.00

Source: Follow-up survey administered 12 months after baseline.

Notes: Results in this table are based on 2,679 – 2,683 respondents who provided valid survey responses to relevant items.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression-adjusted impact estimate (treatment effect).

[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed in the original metric of the outcome variable. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[c] The effect size is the standardized effect size of the difference, which is the "treatment effect" divided by the pooled standard deviation of the treatment and control groups.

[d] This scale averages responses to 3 items (shown in table) on attitudes toward childbearing and the importance of goal setting. Possible values range from 1 to 4 with higher values indicating greater motivation to delay childbearing (alpha coefficient = 0.86).

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

4.4.4 Intentions

RtR had no impact on student intentions to engage in sexual activity. Students in both the treatment and control groups were equal in their expectations of engaging in sexual intercourse or oral sex in the 12 months after the survey. Nearly all of the students reported that they intended to use condoms or birth control if they were to engage in sexual intercourse (Exhibit 4.7).

Site-Level Differences. Site-level analyses revealed that, in BFL, there was a significant favorable program effect on intentions. Students in this site who had participated in *RtR* reported reduced intentions (7.39 percentage point difference) to engage in oral sex in the subsequent 12 months compared with their control group counterparts (Appendix Table A.6).

There were no subgroup differences in the impacts of *RtR* on intentions.

Exhibit 4.7: Short-Term Impacts on Intentions

Outcome	Adjusted Treatment Mean [a]	Unadjusted Control Mean	Treatment Effect [b]	p-value
Intentions				
Percentage of respondents reporting intentions to engage in the following behaviors in the next 12 months:				
Sexual intercourse	52.67	50.69	1.97	0.280
Oral sex	42.41	43.27	-0.86	0.632
Use a condom if they were to have sexual intercourse	91.21	92.11	-0.90	0.403
Use birth control if they were to have sexual intercourse	90.39	89.67	0.72	0.537

Source: Follow-up survey administered 12 months after baseline.

Notes: Results in this table are based on 2,654 – 2,667 respondents who provided valid survey responses to relevant items.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

* p< 0.05, ** p< 0.01, *** p< 0.001.

4.4.5 Skills

RtR had no significant short-term impact on either of two measures of skills: perceived refusal skills and perceived condom negotiation skills (Exhibit 4.8).

Site-Level Differences. Site-level analyses found a modest (SES = .19) but highly significant positive impact on perceived condom negotiation skills in the SDYS sample (Appendix Table A.6). In SDYS, students who had participated in *RtR* reported stronger condom negotiation skills than did students in the control group.

There were no subgroup differences in the impacts of *RtR* on perceived skills.

Exhibit 4.8: Short-Term Impacts on Skills

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	p-value	Effect Size[c]
Skills					
Perceived refusal skills[d]	3.12	3.08	0.04	0.132	0.06
Perceived condom negotiation skills[e]	3.53	3.50	0.03	0.177	0.06

Source: Follow-up survey administered 12 months after baseline.

Notes: Results in this table are based on 2,681–2,685 respondents who provided valid survey responses to relevant items.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression-adjusted impact estimate (treatment effect).

[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed in the original metric of the outcome variable. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[c] The effect size is the standardized effect size of the difference, which is the "treatment effect" divided by the pooled standard deviation of the treatment and control groups.

[d] This scale averages responses to six questions on perceived refusal skills. Possible values range from 1 to 4 with higher values indicating greater perceived skills (alpha coefficient = 0.86).

[e] This scale averages responses to seven questions on perceived condom skills. Possible values range from 1 to 4 with higher values indicating greater perceived skills (alpha coefficient = 0.83).

* p< 0.05, ** p< 0.01, *** p< 0.001.

4.5 Youth Sexual Behavior and Sexual Risk

Despite impacts on some of the intermediate outcomes the model predicts will lead to behavior change, *RtR* did not have an overall impact on either of the two confirmatory outcome measures of current sexual activity and sexual intercourse without birth control (bolded in Exhibit 4.9) at the short-term follow-up (12 months post-baseline). Nor were there statistically significant impacts on reported rates of other sexual behaviors or sexual risk.

Site-level Differences. Site-level analyses revealed a significant and favorable effect of the program on students in the BFL sample (Appendix Table A.7). In BFL, fewer program participants (32.7 percent) engaged in sexual intercourse in the 90 days before the survey compared with their control group counterparts (39.3 percent).

There were no significant subgroup differences in the impacts of *RtR* on sexual behavior or sexual risk.

Exhibit 4.9: Short-Term Impacts on Sexual Behavior and Sexual Risk

Outcome	Adjusted Treatment Mean [a]	Unadjusted Control Mean	Treatment Effect [b]	p-value
Sexual behavior (percentage responding affirmatively)				
Currently sexually active (in last 90 days) [c]	28.02	28.14	-0.11	0.946
Sexual intercourse in the last 90 days	23.66	24.37	-0.72	0.671
Oral sex in the last 90 days	19.24	19.50	-0.26	0.871
Initiation of sexual activity [c, d]	24.98	21.96	-3.02	0.156
Sexual risk (percentage responding affirmatively)				
Sexual intercourse without birth control (in last 90 days)	8.73	8.99	-0.25	0.815
Sexual intercourse without a condom (in last 90 days)	13.57	15.38	-1.81	0.178
Oral sex without a condom (in last 90 days)	16.20	17.33	-1.13	0.444

Source: Follow-up survey administered 12 months after baseline.

Notes: Results in this table are based on 2,661 – 2,667 respondents who provided valid survey responses to relevant items. Confirmatory outcomes are bolded.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[b] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[c] Sexual activity is defined differently across grantees. In Better Family Life, sexual activity refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in LifeWorks or San Diego Youth Services.

[d] The sample size for the initiation of sexual activity outcome is 1836, as this outcome only includes youth who were not sexually active at baseline.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests). (For the two confirmatory outcomes, statistical significance at p<0.05, p<0.01, and p<0.001 implies statistical significance at these levels after applying a Benjamini-Hochberg adjustment for multiple comparisons.)

5. Discussion

This report on short-term findings for the *Reducing the Risk* program model is the first of two impact reports and is not intended to provide comprehensive evidence about the program's effects. Furthermore, this short-term follow-up does not analyze data on important outcomes that reflect the ultimate goals of the Teen Pregnancy Prevention Initiative (prevention of pregnancy and STI transmission). A final assessment of *RtR*'s effectiveness must await the findings from the final, longer-term follow-up survey, conducted two years after this replication began. However, the short-term results presented in this report provide some insight into what we might see at the longer-term. Overall, short-term results are mixed, but some provide grounds for optimism.

RtR was implemented with fidelity in all three replication sites, but attendance varied across the three sites. In BFL and SDYS, a majority of students received at least 75 percent of the classes; in LifeWorks, a little less than half did. One possibility for the lower participation rate in LifeWorks may be the high mobility among these students. Despite variation in attendance, across the three replications, *RtR* succeeded in exposing youth to more information about sexual health and sexual risk prevention than they would otherwise have received.

RtR also demonstrated effectiveness in changing several of the intermediate outcomes, which the logic model predicts will lead to behavior change. When we look at these three replications, it seems that the impact on intermediate outcomes was somewhat stronger than in the original study (Kirby et al., 1991). The original study found positive impacts on knowledge about sexual risk behaviors, which were replicated in the current study. Unlike the earlier study, however, the current study also found impacts on attitudes toward protection.

We found that student attitudes toward use of birth control and condoms were significantly more positive as a result of program participation, an impact that seems to have been driven primarily by the SDYS replication, where students were younger and less likely to have engaged in risky sexual behavior than were students in the other two sites. This suggests that these younger students, who were not yet engaging in risk behaviors, may have been more open to the different and more positive views presented by the program. In contrast, while most students expressed negative attitudes towards risky behavior, we did see some small but potentially adverse program effects on attitudes toward risky sexual behaviors among older students.

Beyond these impacts, the replications showed no significant overall effects on motivation to delay pregnancy or intentions with respect to future sexual behavior. For students in BFL, however, *Reducing the Risk* lowered intentions to have oral sex. Among these same students, though, the program also decreased motivation to delay childbearing.

There were also no overall effects on perceived skills. Given the program's emphasis on developing and practicing skills, this finding was surprising. However, the highly significant positive impact on perceived condom negotiation skills among students in SDYS (who were younger and engaging in fewer risky sexual behaviors at baseline) again suggests a more open attitude towards the curriculum on the part of these students.

The original study found no effects on any sexual risk behavior six months after the program ended, which corresponds roughly to our short-term follow-up at 12 months after baseline. The current study found no impacts on the two confirmatory outcomes (i.e., currently sexually active and sexual intercourse

without birth control). Similarly, no significant impacts were found on any other behavioral outcomes for the overall sample. However, the current study did find a significant favorable program effect on one aspect of sexual behavior (sexual intercourse in the last 90 days) for program participants in one site, BFL. At the short-term follow-up, *RtR* participants in BFL were less likely to have engaged in sexual intercourse in the last 90 days compared with their control group counterparts, an impact which is not seen in the other sites. No other impacts were significantly different in BFL than in the other two sites. However, the overall pattern of impact estimates in BFL was uniformly in the desired direction. Notably, the sample looked different in BFL than in the other two sites. In BFL, a substantially higher proportion had engaged in sexual activity when they entered the study than in the other two sites: almost one-third of the students were sexually active and almost half had some prior sexual experience. The racial composition of the sample was also much different in BFL than in the other sites, with a much higher proportion of Black youth. The longer-term report will provide further insight on whether and how these differences influence program effectiveness.

In addition to the favorable short-term findings on some intermediate outcomes and on one aspect of sexual behavior in one site, there is another reason for optimism about potential longer-term effects of the *Reducing the Risk* program model. Students in the original study sample were considerably older, on average, than the students in this study; three-quarters of students in the original study were in 10th grade or higher, and the remaining one-quarter were in 9th grade. By contrast, students in the sample for this study were predominantly 9th and 10th graders. In SDYS, the students were even younger, with a substantial percentage of 8th graders. In that site, less than 10 percent of students were sexually active when they entered the study, compared with the 37 percent of students who entered the original study. It seems plausible that, as youth mature and more of them become sexually experienced, we may detect additional impacts on sexual behavior. Our findings on the intermediate outcomes suggest that the younger, less sexually experienced students in SDYS were more responsive to program messages; the improved attitudes and skills among *RtR* participants in this site may translate into reduced sexual risk behavior over time.

The TPP Replication Study was designed to address important research and policy questions about the effectiveness of evidence-based programs, and what happens when they are taken to scale, replicated with different populations, and in different settings. The three replicated program models were intentionally selected to maximize what could be learned about different strategies and to begin to address identified gaps in the teen pregnancy prevention research. This report, part of a larger set of reports on replications of evidence-based program models, provides important information on the early effectiveness of *Reducing the Risk.*

While we found no impact on the two confirmatory outcomes and other findings were fixed, at the very least, these short-term findings suggest that strong replications of the program can have impacts on some intermediate outcomes, thought to lead to behavior change, that are comparable to and even exceeding those reported by the evaluator and program developer in the original study.

Short-term impact reports on the other two models (*¡Cuídate!* and *Safer Sex Intervention*) will shed additional light on our understanding of strategies for addressing youth risk behavior and promoting healthy choices for youth. The three final reports on longer-term outcomes will provide more comprehensive evidence on the effectiveness of these programs on sexual risk-taking behaviors and their consequences.

References

Constantine, J., Player D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). *An evaluation of teachers trained through different routes to certification, final report* (NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gleason, P., Clark, M., Tuttle, C. & Dwyer E. (2010). *The Evaluation of Charter School Impacts: Executive Summary* (NCEE 2010-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goesling, B., Colman, C., Trenholm, C., Terzian, M. & Moore, K. (2014). *Programs to reduce teen pregnancy, sexually transmitted infections, and associated risk behaviors: A systematic review*. Journal of Adolescent Health, 54(5), 499-507.

Kappeler, E. and Farb, A. (2014). *Historical Context for the Creation of the Office of Adolescent Health and the Teen Pregnancy Prevention Program.* Journal of Adolescent Health, 54(3S), S3-S9.

Kirby, D., Barch, R.P., Leland, N., & Fetro, J.V. (1991) Reducing the Risk: impact of a new curriculum on sexual risk-taking. *Family Planning Perspectives*, 23 (6) 253-263.

Lezin, N., Rolleri, L.A., Wilson, M.M., Fuller, T.R., & Firpo-Triplett, R. (2010). *Reducing the Risk* Adaptation Kit. Santa Cruz, CA: ETR Associates.

Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods.* Thousand Oaks, CA: Sage Publications.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schochet, P.Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics*, 33(1), 62-87.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014). Procedures and Standards Handbook, version 3.0 Retrieved from http://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

U.S. Department of Health and Human Services. (2015). HHS Teen Pregnancy Prevention Evidence Review [website]. http://tppevidencereview.aspe.hhs.gov/

U.S. Department of Health and Human Services. (2010). Teenage pregnancy prevention: Replication of evidence-based programs. Funding opportunity announcement and application instructions. Office of Adolescent Health, Office of Public Health and Science.

Appendix A: Site-Level Impacts

This study was carefully designed such that when data from all three replication sites were pooled into a single analysis, the combined sample would be large enough for the study to be adequately powered to detect effects of the *RtR* intervention on all of the outcomes of interest. Although the pooled analysis is the primary focus of this study, there was clearly considerable interest on the part of study stakeholders in examining the results from each of the three replication sites, and the large sample sizes preserve the ability to conduct these analyses. Therefore this appendix presents site-specific impact estimates for each of the outcomes reported in the main text. We urge two major types of caution for readers who examine the results from the individual sites. The first is that the study was not designed to have large enough sample sizes in each individual site to have a good chance of detecting a treatment effect for all of the outcomes of interest. Thus, in a single site, lack of statistical significance could be the result of either an insufficiently large sample to detect a true effect, or it could mean that the intervention did not produce an effect on the outcome. Second, there are a large number of results presented in Appendix A, and these results are not adjusted for multiple comparisons. Some statistically significant findings would be expected purely by chance among such a large number of tests. Therefore, the findings in these tables should be interpreted with caution. The final column of each table shows the statistical result for a test of differences in the treatment effect across sites. When a statistically significant difference is found, the corresponding site-specific impacts are discussed in the main text, as we only interpret site-specific impacts when a significant difference between sites is found.

A.1 Impacts on Exposure to Program Information, by Site

Outcome	Better Family Life (n=941)				LifeWorks (n=853)				San Diego Youth Services (n= 895)				p-value for the Test of Differences Across Sites[a]
	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	
Percentage of respondents that reported receiving information on the following topics[d]:													
Relationships or marriage	81.01	81.84	-0.83	0.777	80.53	80.78	-0.25	0.934	80.94	78.44	2.50	0.441	0.726
Abstinence from sex	77.15	72.09	5.06	0.109	72.92	62.29	10.63 **	0.001	74.16	67.62	6.54	0.057	0.457
Birth control methods	69.56	60.60	8.96 **	0.008	78.88	65.27	13.61 ***	0.000	68.65	63.12	5.53	0.143	0.293
Where to obtain birth control	68.78	59.89	8.89 **	0.008	72.39	58.92	13.47 ***	0.000	69.24	54.81	14.43 ***	0.000	0.473
Sexually transmitted infections	88.87	87.53	1.34	0.615	81.91	81.51	0.40	0.886	82.20	76.36	5.84 *	0.045	0.355
How to talk with partner about sex and birth control	84.54	76.15	8.39 **	0.006	77.09	62.59	14.50 ***	0.000	69.34	58.44	10.90 ***	0.001	0.375
How to say no to sex	85.53	78.59	6.94 *	0.019	79.89	66.59	13.30 ***	0.000	82.27	71.69	10.58 **	0.001	0.325
How babies are made	90.59	87.53	3.06	0.166	84.41	84.43	-0.02	0.993	88.83	88.05	0.78	0.736	0.601

Source: Follow-up survey administered six months after baseline.

[a] This column shows the results for statistical tests of whether the treatment effect varies among the three sites.

[b] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[c] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[d] Refers to information received in the 12 months prior to the survey administration.

p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

A.2 Impacts on Knowledge of Pregnancy Risk and STI Risk, by Site

Outcome	Better Family Life (n=941)				LifeWorks (n=853)				San Diego Youth Services (n= 895)				p-value for the Test of Differences Across Sites [a]
	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	
Knowledge of pregnancy risk (percent of items respondents answered correctly) [d]	59.34	55.96	3.38	0.062	71.91	69.53	2.38	0.204	64.70	58.38	6.32 ***	0.001	0.303
Percentage of respondents correctly answering each item:													
Used correctly, how much can birth control pills reduce pregnancy risk?	50.50	44.17	6.33	0.051	67.02	65.21	1.81	0.591	58.27	54.03	4.24	0.213	0.626
Used correctly, how much can condoms reduce pregnancy risk?	54.96	55.01	-0.05	0.988	70.42	67.40	3.02	0.348	64.50	56.10	8.40 **	0.009	0.165
A couple that has had unprotected sex and not gotten pregnant does not have to worry about getting pregnant.	76.91	76.15	0.76	0.768	84.29	85.16	-0.87	0.743	80.06	78.18	1.88	0.481	0.764
A woman is protected from pregnancy the day she begins taking the pill.	55.16	48.51	6.65	0.059	65.97	60.34	5.63	0.126	56.67	45.19	11.48 **	0.003	0.503
Knowledge of STI risk (percent of items respondents answered correctly) [e]	59.15	54.52	4.63 ***	0.001	63.75	60.46	3.29 *	0.018	58.14	53.29	4.85 ***	0.001	0.687
Percentage of respondents correctly answering each item:													
Once you are infected with HIV you are infected for life	72.05	67.75	4.30	0.142	79.71	76.16	3.55	0.241	78.60	73.25	5.35	0.079	0.915
There is a vaccine to prevent girls from getting HPV	41.68	34.15	7.53 *	0.022	45.69	46.96	-1.27	0.709	48.08	43.12	4.96	0.148	0.165
All STD/STIs can be cured by taking medicine	62.53	55.28	7.25 *	0.019	71.84	69.83	2.01	0.529	64.67	60.26	4.41	0.170	0.498
A person with an STD/STI who looks and feels healthy cannot transmit the infection to others	76.80	73.17	3.63	0.172	83.15	82.73	0.42	0.880	77.44	73.51	3.93	0.155	0.604
Some STDs/STIs put you at greater risk of HIV	66.67	65.31	1.36	0.710	59.63	62.29	-2.66	0.488	57.36	58.70	-1.34	0.741	0.740
About 1 out of 4 sexually active teens gets an STD/STI every year	68.42	66.12	2.30	0.480	60.75	54.74	6.01	0.075	61.74	57.66	4.08	0.228	0.731
You can get an STD/STI from having oral sex	78.86	72.90	5.96	0.064	74.62	70.80	3.82	0.258	61.67	56.10	5.57	0.115	0.890
Used correctly, how much can condoms decrease the risk of HIV?	45.90	41.46	4.44	0.172	58.14	50.85	7.29 *	0.030	50.95	48.83	2.12	0.530	0.555
You can get HIV even if you unprotected sex only 1 or 2 times	60.19	56.91	3.28	0.299	68.84	66.91	1.93	0.554	65.29	59.22	6.07	0.065	0.660
Used correctly, how much can condoms decrease the risk of gonorrhea?	36.23	31.71	4.52	0.162	44.41	37.23	7.18 *	0.033	41.05	29.87	11.18 **	0.001	0.367
Used correctly, how much can birth control pills decrease the risk of HIV?	49.68	44.72	4.96	0.105	59.03	56.20	2.83	0.371	47.03	43.12	3.91	0.219	0.890

Outcome	Better Family Life (n=941)				LifeWorks (n=853)				San Diego Youth Services (n= 895)				p-value for the Test of Differences Across Sites [a]
	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	
Used correctly, how much can birth control pills decrease the risk of gonorrhea	50.48	44.72	5.76	0.062	59.41	50.85	8.56 **	0.008	43.51	35.84	7.67 *	0.019	0.813

Source: Follow-up survey administered 12 months after baseline.

[a] This column shows the results for statistical tests of whether the treatment effect varies among the three sites.

[b] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[c] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[d] Score based on the four items below. Values represent the average percent of items answered correctly by respondent for each group.

[e] Score based on the 12 items below. Values shown represent the average percent of items answered correctly by respondent for each group.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

A.3 Impacts on Attitudes Toward Protection, by Site

Outcome	Better Family Life (n=940)					LifeWorks (n=853)					San Diego Youth Services (n= 895)					p-value for the Test of Differences Across Sites [a]
	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	SES[d]	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	SES[d]	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	SES[d]	
Attitudes toward protection [e]	3.20	3.19	0.01	0.605	0.03	3.18	3.14	0.04	0.144	0.09	3.17	3.05	0.12 ***	0.000	0.30	0.006 **
Birth control pills should always be used if a person your age has sexual intercourse	3.22	3.30	-0.07	0.202	-0.09	3.28	3.23	0.06	0.342	0.07	3.20	3.13	0.07	0.241	0.09	0.156
Birth control is too much trouble to use (reverse)	3.19	3.24	-0.05	0.296	-0.07	3.06	3.11	-0.05	0.330	-0.07	3.00	2.94	0.06	0.245	0.08	0.210
Birth control is pretty easy to get	3.01	2.94	0.07	0.215	0.08	2.78	2.63	0.15 **	0.007	0.18	2.83	2.67	0.16 **	0.004	0.20	0.406
Birth control is important to make sex safer	3.22	3.18	0.04	0.398	0.06	3.32	3.26	0.06	0.257	0.08	3.23	3.10	0.13 *	0.016	0.17	0.467
Birth control has too many side effects (reverse)	2.44	2.45	-0.01	0.829	-0.02	2.71	2.61	0.09	0.100	0.12	2.67	2.53	0.14 *	0.020	0.18	0.153
Using birth control is morally wrong (reverse)	3.23	3.21	0.02	0.691	0.03	3.23	3.28	-0.05	0.274	-0.07	3.19	3.04	0.16 **	0.002	0.21	0.011 *
Condoms are too much trouble to use (reverse)	3.41	3.42	-0.01	0.899	-0.01	3.25	3.24	0.01	0.831	0.01	3.30	3.17	0.13 *	0.015	0.17	0.141
Condoms are pretty easy to get	3.34	3.32	0.02	0.690	0.03	3.34	3.24	0.09	0.077	0.12	3.32	3.17	0.15 **	0.005	0.19	0.211
Condoms are important to make sex safer	3.67	3.72	-0.05	0.281	-0.08	3.64	3.60	0.03	0.499	0.05	3.58	3.53	0.05	0.361	0.07	0.303
Using condoms means you don't trust your partner (reverse)	3.35	3.26	0.09	0.069	0.12	3.41	3.43	-0.02	0.771	-0.02	3.36	3.21	0.15 **	0.004	0.19	0.077
Using condoms is morally wrong (reverse)	3.52	3.54	-0.02	0.682	-0.03	3.51	3.50	0.01	0.808	0.02	3.56	3.38	0.18 ***	0.000	0.26	0.005 **
Condoms decrease sexual pleasure (reverse)	2.72	2.64	0.08	0.150	0.09	2.56	2.55	0.00	0.962	0.00	2.81	2.67	0.14 *	0.013	0.16	0.221

Source: Follow-up survey administered 12 months after baseline.

[a] This column shows the results for statistical tests of whether the treatment effect varies among the three sites.

[b] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[c] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed in the original metric of the outcome variable. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[d] The effect size is the standardized effect size of the difference, which is the "treatment effect" divided by the pooled standard deviation of the treatment and control groups.

[e] This construct averages responses to 12 items (shown in table) on attitudes towards condoms and birth control. Possible values range from 1 to 4 with higher values indicating more positive attitudes toward protection.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

A.4 Impacts on Attitudes Toward Risky Sexual Behavior, by Site

Outcome	Better Family Life (n=938)				LifeWorks (n=848)				San Diego Youth Services (n= 889)				p-value for the Test of Differences Across Sites[a]
	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	
Attitudes toward risky behavior (percent of items respondents agreed with) [d]	3.88	2.80	1.08	0.259	7.13	5.97	1.16	0.246	4.77	4.65	0.12	0.903	0.713
Percentage of respondents agreeing with each item:													
It's OK to have sex with someone on your first date	4.79	5.71	-0.92	0.584	10.85	8.80	2.05	0.241	5.90	4.99	0.91	0.608	0.464
It's OK to have sex with someone the same night you meet them	4.21	4.08	0.13	0.932	8.56	8.31	0.25	0.875	5.53	4.20	1.33	0.411	0.844
It's OK to have sex with several different people in the same month	2.82	2.17	0.65	0.662	7.12	7.09	0.03	0.983	3.96	5.51	-1.55	0.330	0.587
It's okay to have sex without protection	3.45	2.45	1.00	0.432	5.23	2.69	2.54	0.056	4.79	4.46	0.33	0.804	0.483
It's OK to have sex with someone when you know they are someone else's girlfriend/boyfriend	5.37	2.45	2.92 *	0.032	4.77	3.67	1.10	0.436	3.69	3.67	0.02	0.987	0.331
It's OK to have sex with someone if you are drunk or high	4.12	2.45	1.67	0.272	7.91	7.09	0.82	0.607	5.40	5.51	-0.11	0.945	0.723
It's OK to have sex with someone if you know they are drunk or high	2.44	0.27	2.17	0.076	5.53	4.16	1.37	0.281	4.25	4.20	0.05	0.966	0.484

Source: Follow-up survey administered 12 months after baseline.

[a] This column shows the results for statistical tests of whether the treatment effect varies among the three sites.

[b] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[c] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[d] Score based on the seven items (shown below) represents the average percent of items agreed with by respondent for each group.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

A.5 Impacts on Motivation to Delay Childbearing, by Site

Outcome	*Better Family Life (n=938)*					*LifeWorks (n=853)*					*San Diego Youth Services (n= 892)*					*p-value for the Test of Differences Across Sites*[a]
	Adj. T Mean[b]	*Unadj. C Mean*	*T Effect*[c]	*p-value*	*SES*[d]	*Adj. T Mean*[b]	*Unadj. C Mean*	*T Effect*[c]	*p-value*	*SES*[d]	*Adj. T Mean*[b]	*Unadj. C Mean*	*T Effect*[c]	*p-value*	*SES*[d]	
Motivation to delay childbearing [e]	*3.64*	*3.75*	*-0.10 **	*0.014*	*-0.18*	*3.71*	*3.66*	*0.04*	*0.354*	*0.07*	*3.70*	*3.65*	*0.05*	*0.256*	*0.09*	*0.018 **
You have goals you want to accomplish before having a child	3.59	3.71	-0.12 *	0.017	-0.18	3.66	3.64	0.02	0.749	0.02	3.65	3.60	0.05	0.362	0.07	0.049 *
It is important for you to finish school before you have a child	3.66	3.77	-0.11 *	0.018	-0.17	3.72	3.65	0.07	0.148	0.11	3.72	3.68	0.04	0.414	0.06	0.016 *
It is important to have a job and a stable income before you have a child	3.69	3.77	-0. 08	0.067	-0.13	3.73	3.70	0.03	0.544	0.05	3.73	3.67	0.07	0.181	0.11	0.064

Source: Follow-up survey administered 12 months after baseline.

[a] This column shows the results for statistical tests of whether the treatment effect varies among the three sites.

[b] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[c] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed in the original metric of the outcome variable. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[d] The effect size is the standardized effect size of the difference, which is the "treatment effect" divided by the pooled standard deviation of the treatment and control groups.

[e] This scale averages responses to 3 items (shown in table) on attitudes toward childbearing and the importance of goal setting. Possible values range from 1 to 4 with higher values indicating greater motivation to delay childbearing.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

A.6 Impacts on Intentions and Skills, by Site

Outcome	Better Family Life (n=938)					LifeWorks (n=853)					San Diego Youth Services (n= 894)					p-value for the Test of Differences Across Sites[a]
	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	SES[d]	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	SES[d]	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	SES[d]	
Intentions																
Percentage of respondents reporting intentions to engage in the following behaviors in the next 12 months:																
Sexual intercourse	58.96	62.64	-3.68	0.220		62.29	57.25	5.04	0.108		37.72	32.38	5.34	0.103		0.062
Oral sex	38.89	46.28	-7.39 *	0.011		53.95	52.58	1.37	0.650		34.85	30.45	4.40	0.158		0.015 *
Use a condom if they were to have sexual intercourse	91.81	93.70	-1.89	0.300		89.64	92.89	-3.25	0.084		92.30	89.76	2.54	0.180		0.076
Use birth control if they were to have sexual intercourse	89.64	90.63	-0.99	0.618		90.20	88.73	1.47	0.473		91.58	89.76	1.82	0.378		0.560
Skills																
Perceived refusal skills (scale score)[e]	3.10	3.09	0.02	0.709	0.02	3.14	3.05	0.09	0.062	0.12	3.12	3.10	0.02	0.700	0.03	0.472
Perceived condom negotiation skills (scale score) [f]	3.61	3.66	-0.05	0.182	-0.09	3.52	3.48	0.04	0.234	0.08	3.47	3.37	0.10 **	0.008	0.19	0.015 *

Source: Follow-up survey administered 12 months after baseline.

[a] This column shows the statistical result for the test of differences in the treatment effect across sites.

[b] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[c] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. For outcomes reported as percentages, the treatment effect is expressed as a difference in percentage points. For scale outcomes, the treatment effect is expressed in the original metric of the outcome variable. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[d] The effect size is the standardized effect size of the difference, which is the "treatment effect" divided by the pooled standard deviation of the treatment and control groups.

[e] This scale averages responses to 6 questions on perceived refusal skills. Possible values range from 1 to 4 with higher values indicating greater perceived skills.

[f] This scale averages responses to 7 questions on perceived condom skills. Possible values range from 1 to 4 with higher values indicating greater perceived skills.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

A.7 Impacts on Sexual Behavior and Sexual Risk, by Site

Outcome	Better Family Life (n=934)				LifeWorks (n=848)				San Diego Youth Services (n= 885)				p-value for the Test of Differences Across Sites[a]
	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	Adj. T Mean[b]	Unadj. C Mean	T Effect[c]	p-value	
Sexual behavior (percentage responding affirmatively)													
Currently sexually active (in last 90 days)[d]	38.35	41.37	-3.02	0.285	35.29	31.13	4.16	0.160	10.99	12.30	-1.31	0.672	0.193
Sexual intercourse in the last 90 days	32.71	39.34	-6.63 *	0.015	30.87	25.74	5.13	0.073	8.65	8.62	0.03	0.992	0.011 *
Oral sex in the last 90 days	23.06	23.29	-0.23	0.932	26.82	25.55	1.27	0.647	7.44	9.42	-1.98	0.497	0.722
Sexual risk (percentage responding affirmatively)													
Sexual intercourse without birth control (in last 90 days)	11.20	13.11	-1.91	0.300	12.33	11.03	1.30	0.493	2.83	2.87	-0.04	0.984	0.475
Sexual intercourse without a condom (in last 90 days)	17.47	22.13	-4.66 *	0.037	19.65	19.85	-0.20	0.933	3.93	4.18	-0.25	0.918	0.282
Oral sex without a condom (in last 90 days)	17.89	19.45	-1.56	0.529	23.97	23.83	0.14	0.956	6.35	8.38	-2.03	0.455	0.827

Source: Follow-up survey administered 12 months after baseline.

[a] This column shows the results for statistical tests of whether the treatment effect varies among the three sites.

[b] The treatment group mean is regression-adjusted, calculated as the sum of the unadjusted control group mean and the regression adjusted impact estimate (treatment effect).

[c] The treatment effect was estimated in a multi-level model that controls for site/school/semester/gender blocks and other covariates. The treatment effect is expressed as a difference in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[d] Sexual activity is defined differently across grantees. In Better Family Life, sexual activity refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in LifeWorks or San Diego Youth Services.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

Appendix B: Subgroup Impacts

B.1 Impacts on Exposure to Program Information at Short-term Follow-up, by Subgroup

	Treatment Effect [a]	p-value [b]
Received information about abstinence (percentage of respondents)		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n =1,848)	9.83***	0.000
Ever sexually active at Baseline (n =837)	2.10	0.511
Received information about birth control methods (percentage of respondents)		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n =1,846)	12.95***	0.000
Ever sexually active at Baseline (n =836)	2.30	0.488
Received information about where to obtain birth control (percentage of respondents)		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n =1,848)	16.08***	0.000
Ever sexually active at Baseline (n =836)	3.64	0.265
Subgroup: Respondent age		
Respondent less than age 15 (n =1,548)	15.22***	0.000
Respondent age 15 or older (n =1,136)	7.85**	0.006
Received information about talking to partner about having sex or using birth control (percentage of respondents)		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n =1,851)	14.75***	0.000
Ever sexually active at Baseline (n =836)	3.54	0.254
Received information about how to say no to sex (percentage of respondents)		
Subgroup: Respondent gender		
Male (n =1,354)	15.34***	0.000
Female (n =1,333)	4.81*	0.048
Received information about how babies are made (percentage of respondents)		
Subgroup: Respondent Race		
Hispanic (n =1,177)	-1.23	0.528
Black (n =965)	4.14	0.056
White (n =288)	-4.49	0.247
Other (n =258)	9.41*	0.023

Source: Follow-up survey administered 12 months after baseline.

Notes: Impact estimates for subgroups are shown only if a test for differences in impacts among the subgroups met the study criterion for statistical significance (p<0.05). For example, a test result indicated that the treatment effect on learning about where to obtain birth control was significantly different for the sexually inexperienced at baseline subgroup.

[a] This column shows the estimated treatment effect (treatment/control differences in the percent reporting receiving information) for the subgroup indicated in the row

[b] This column shows the statistical test result for whether the treatment effect for the subgroup indicated in the row was significantly different than zero.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

B.2 Impacts on Attitudes toward Risky Behavior, by Subgroup

	Treatment Effect [a]	p-value [b]
Subgroup: Respondent age		
Respondent less than age 15 (n =1,545)	-0.30	0.690
Respondent age 15 or older (n =1,130)	2.31**	0.008
Subgroup: Respondent Race		
Hispanic (n =1,170)	-0.71	0.401
Black (n =960)	1.18	0.214
White (n =287)	5.66***	0.001
Other (n =258)	0.66	0.717

Source: Follow-up survey administered 12 months after baseline.

Notes: Impact estimates for subgroups are shown only if a test for differences in impacts among the subgroups met the study criterion for statistical significance (p<0.05). For example, a test result indicated that the treatment effect on attitudes toward risky behavior was significantly different for older respondents.

[a] This column shows the estimated treatment effect (treatment/control difference in the percent of items respondent agreed with) for the subgroup indicated in the row.

[b] This column shows the statistical test result for whether the treatment effect for the subgroup indicated in the row was significantly different than zero.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

B.3 Impacts on Motivation to Delay Childbearing, by Subgroup

	Treatment Effect [a]	p-value [b]
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n =1,847)	-0.05	0.142
Ever sexually active at Baseline (n =836)	0.07	0.112

Source: Follow-up survey administered 12 months after baseline.

Notes: Impact estimates for subgroups are shown only if a test for differences in impacts among the subgroups met the study criterion for statistical significance (p<0.05). For example, a test result indicated that the treatment effect on motivation to delay childbearing was significantly different for the sexually inexperienced at baseline subgroup.

[a] This column shows the estimated treatment effect (treatment/control difference in the percent of items respondent agreed with) for the subgroup indicated in the row.

[b] This column shows the statistical test result for whether the treatment effect for the subgroup indicated in the row was significantly different than zero.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).

Appendix C: Measures

The measures we used to examine short-term program impacts stem from our research questions (Section 3.1) and logic model (Exhibit 2.2) and are organized into three categories:

- Exposure to program information,
- Intermediate outcomes, and
- Youth sexual behavior.

Measures in the first category (*exposure to program information*) reflect receipt of sexuality education and reproductive health information. These provide insight into *RtR's* success in reaching youth. Measures of *intermediate outcomes* indicate the extent to which youth assimilated the program's messages and reflected them in their knowledge, attitudes, motivation, intentions, and skills—all of which are hypothesized precursors of change in youth's sexual behavior. Measures of *youth sexual behavior* include measures of sexual activity and sexual risk behavior (e.g., unprotected sexual activity). In the sections that follow, we describe each category by defining constituent measures and their construction.

C.1 Exposure to Program Information

To assess whether *RtR* increased exposure to information on sexual health, contraception, and STI transmission and prevention, at the short-term follow-up, we asked youth about their receipt of sexuality education and reproductive health information. [21] On the survey, they responded to a series of questions asking about their exposure to information about: (a) relationships or marriage; (b) abstinence from sex; (c) birth control methods; (d) where to obtain birth control; (e) STIs; (f) how to talk with a partner about sex and birth control; (g) how to say no to sex; and (h) how babies are made. For each, youth were asked whether they had "received information or learned about" the topic in the 12 months prior to survey administration. Responses were coded in a binary fashion, as 1 = "yes" and 0 = "no."

C.2 Intermediate Outcomes

Intermediate outcomes are those expected to portend changes in behavior. At the short-term follow-up, we asked youth a wide variety of questions to gauge their understanding, thoughts, beliefs, and perceptions of topics addressed by the program. We organized these measures conceptually into five domains: knowledge, attitudes, motivation, intentions, and skills. Using survey items relevant to each domain, we conducted factor analyses and reliability testing to construct composite measures in each domain, where this was possible. In addition, we used baseline data (when the same items were asked) to examine the stability over time of composite measures, and examined the follow-up data by racial-ethnic subgroup to assess the stability of constructs.

Knowledge

To examine program-related changes in youth's sexual health knowledge, we constructed two measures: *knowledge of pregnancy risk* and *knowledge of STI risk*. These measures were defined conceptually and

[21] At baseline, before random assignment, youth were asked these same questions about the 12 month period preceding the study. The reference period for the follow-up survey included the period in which treatment group members were offered the intervention (and controls were not).

constructed to differentiate accurate knowledge from misinformation. They may be considered tests of understanding of the factors contributing to pregnancy and STIs. The construction of these measures is described below and detailed information about their component items is presented in Exhibit C2.1.

- ***Knowledge of pregnancy risk*** is a composite measure that is the mean (multiplied by 100) of four binary variables regarding knowledge of the extent to which contraceptive methods can prevent pregnancy and circumstances under which pregnancy is possible (See Exhibit C2.1 for coding and other details). Scores on this scale range from 0 to 100 and represent the percentage of correct answers across the four items. Higher values indicate more accurate knowledge.
- ***Knowledge of STI risk*** is a composite measure that is the mean of 12 binary variables (multiplied by 100) pertaining to knowledge of STI prevention, transmission, and treatment (See Exhibit C2.1 for coding and other details.) Scores on this scale range from 0 to 100 and represent the percentage of correct answers across the 12 items. Higher values indicate more accurate knowledge.

Exhibit C2.1: Knowledge Scales and Component Items

Component Items	Coding
Knowledge of Pregnancy Risk (4 items)	
Used correctly, how much can birth control pills reduce pregnancy risk?	Youth responded to this question on a scale from 1= "Not at all" to 4="Completely." This item was recoded into a binary variable where the correct response ("a lot") was coded as 1 and all other responses were coded as 0.
Used correctly, how much can condoms reduce pregnancy risk?	Youth responded to this question on a scale from 1= "Not at all" to 4="Completely." This item was recoded into a binary variable where the correct response ("a lot") was coded as 1 and all other responses were coded as 0.
A couple that has had unprotected sex and not gotten pregnant does not have to worry about getting pregnant.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was false, and 0 indicates they were sure or thought the statement was true or did not know.
A woman is protected from pregnancy the day she begins taking the pill.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was false and 0 indicates they were sure or thought the statement was true or did not know.
Knowledge of STI Risk (12 items)	
You can't get infected with HIV if you have sex only once or twice w/o a condom.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was false and 0 indicates they were sure or thought the statement was true or did not know.
Once you are infected with HIV you are infected for life.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was true and 0 indicates they were sure or thought the statement was false or did not know.
There is a vaccine to prevent girls from getting HPV.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was true and 0 indicates they were sure or thought the statement was false or did not know.
All STDs/STIs can be cured by taking medicine.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was false and 0 indicates they were sure or thought the statement was true or did not know.
A person with an STD/STI who looks and feels healthy cannot transmit the infection to others.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was false and 0 indicates they were sure or thought the statement was true or did not know.
Some STDs/STIs put you at greater risk of HIV.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was true and 0 indicates they were sure or thought the statement was false or did not know.
About 1 out of 4 sexually active teens gets an STD/STI every year.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was true and 0 indicates they were sure or thought the statement was false or did not know.

Component Items	Coding
You can get an STD/STI from having oral sex.	Youth indicated the veracity of this statement, responding on a scale from 1= "I am sure it's true" to 5 = "I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was true and 0 indicates they were sure or thought the statement was false or did not know.
Used correctly, how much can condoms decrease the risk of HIV?	Youth responded to this question on a scale from 1 = "Not at all" to 4 = "Completely." This item was recoded into a binary variable where the correct response ("a lot") was coded as 1 and all other responses were coded as 0.
Used correctly, how much can condoms decrease the risk of gonorrhea?	Youth responded to this question on a scale from 1 = "Not at all" to 4 = "Completely." This item was recoded into a binary variable where the correct response ("a lot") was coded as 1 and all other responses were coded as 0.
Used correctly, how much can birth control pills decrease the risk of HIV?	Youth responded to this question on a scale from 1 = "Not at all" to 4 = "Completely." This item was recoded into a binary variable where the correct response ("not at all") was coded as 1 and all other responses were coded as 0.
Used correctly, how much can birth control pills decrease the risk of gonorrhea?	Youth responded to this question on a scale from 1 = "Not at all" to 4 = "Completely." This item was recoded into a binary variable where the correct response ("not at all") was coded as 1 and all other responses were coded as 0.

Attitudes

The short-term survey included 24 items querying attitudes toward sexual behaviors, sexual risks, and contraceptive methods. From among these, we constructed two measures to examine program impacts on youths' sexual health attitudes: *attitudes toward protection* and *attitudes toward risky behavior*. These measures are described below and detailed information about their component items is presented in Exhibit C2.2.

- ***Attitudes toward protection*** is a composite measure that is the mean of responses to 12 items about the importance of using condoms and/or birth control during sexual activity. (See Exhibit C2.2 for coding and other details.) Scores on this scale represent the level of support for using protection. They range from 1 to 4 with high scores indicating positive and supportive attitudes toward contraceptive use to prevent STIs and/or pregnancy. The measure demonstrates acceptable internal consistency reliability ($\alpha = 0.75$). [22]

- ***Attitudes toward risky behavior*** is a composite measure that is the mean of seven binary items (multiplied by 100) querying the acceptability and normativeness of risky sexual behaviors. (See Exhibit C2.2 for coding and other details.) Scores on this scale range from 0 to 100 and represent the percent of items agreed with: Higher values reflect more support for risky behavior. The measure demonstrates good internal consistency reliability ($\alpha = 0.82$).

[22] As a general rule of thumb, the internal consistency of scales with reliability coefficients between 0.70 – 0.79 is considered "acceptable," between 0.80 – 0.89 is considered "good," and 0.90 or greater is considered "excellent."

Exhibit C2.2: Attitudes Scales and Component Items

Component Items	Coding
Attitudes Toward Protection (12 items)	
Birth control pills should always be used if a person your age has sexual intercourse.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward birth control.
Birth control is too much trouble to use.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." High values indicate more positive attitudes toward birth control.
Birth control is pretty easy to get.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward birth control.
Birth control is important to make sex safer.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward birth control.
Birth control has too many side effects.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." High values indicate more positive attitudes toward birth control.
Using birth control is morally wrong.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." High values indicate more positive attitudes toward birth control.
Condoms are too much trouble to use.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." High values indicate more positive attitudes toward condoms.
Condoms are pretty easy to get.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward condoms.
Condoms are important to make sex safer.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward condoms.
Using condoms means you don't trust your partner.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." High values indicate more positive attitudes toward condoms.
Using condoms is morally wrong.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." High values indicate more positive attitudes toward condoms.
Condoms decrease sexual pleasure.	Youth expressed their agreement with this statement, responding on a scale from 1= "Strongly agree" to 4 = "Strongly disagree." High values indicate more positive attitudes toward condoms.
Attitudes Toward Risky Behavior (7 items)	
It's OK to have sex with someone on your first date.	Youth expressed their agreement with this statement by selecting it if it reflected their views on engaging in sex. Responses were coded in a binary fashion, as 1 when the statement was selected and 0 when not selected.
It's OK to have sex with someone the same night you meet them.	Youth expressed their agreement with this statement by selecting it if it reflected their views on engaging in sex. Responses were coded in a binary fashion, as 1 when the statement was selected and 0 when not selected.

Component Items	Coding
It's OK to have sex with several different people in the same month.	Youth expressed their agreement with this statement by selecting it if it reflected their views on engaging in sex. Responses were coded in a binary fashion, as 1 when the statement was selected and 0 when not selected.
It's OK to have sex without protection.	Youth expressed their agreement with this statement by selecting it if it reflected their views on engaging in sex. Responses were coded in a binary fashion, as 1 when the statement was selected and 0 when not selected.
It's OK to have sex with someone when you know they are someone else's girlfriend/boyfriend.	Youth expressed their agreement with this statement by selecting it if it reflected their views on engaging in sex. Responses were coded in a binary fashion, as 1 when the statement was selected and 0 when not selected.
It's OK to have sex with someone if you are drunk or high.	Youth expressed their agreement with this statement by selecting it if it reflected their views on engaging in sex. Responses were coded in a binary fashion, as 1 when the statement was selected and 0 when not selected.
It's OK to have sex with someone if you know they are drunk or high.	Youth expressed their agreement with this statement by selecting it if it reflected their views on engaging in sex. Responses were coded in a binary fashion, as 1 when the statement was selected and 0 when not selected.

Motivation

The short-term survey included 22 items related to youth's motivation to engage in safe sexual practices and reduce their risk. From these, we developed a measure of motivation to delay childbearing. It is the average of three items related to reasons for delaying childbearing (See Exhibit C2.3 for coding and other details.) Scores on this scale range from 1 to 4 with higher scores indicating more motivation to wait to have a child. The scale demonstrated good internal consistency reliability ($\alpha = 0.86$).

Exhibit C2.3: Motivation Scale and Component Items

Component Items	Coding
Motivation to Delay Childbearing (3 items)	
You have goals you want to accomplish before having a child.	Youth responded to this question on a scale from 1 = "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more agreement.
It is important for you to finish school before you have a child.	Youth responded to this question on a scale from 1 = "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more agreement.
It is important to have a job and a stable income before you have a child.	Youth responded to this question on a scale from 1 = "Strongly agree" to 4 = "Strongly disagree." We reverse coded this item so that higher values indicate more agreement.

Intentions

We used the four items presented in Exhibit C2.4 to examine impacts on youth's intended or anticipated sexual behavior in the coming year.

Exhibit C2.4: Intentions Measures

Item	Coding
Do you intend to have sexual intercourse in the next year, if you have the chance?	Youth responded to this question on a scale from 1 = "Yes, definitely" to 4 = "No, definitely not." This item was recoded into a binary variable where affirmative responses (definitely, probably) were coded as 1 and negative responses (definitely not, probably not) were coded as 0.
Do you intend to have oral sex in the next year, if you have the chance?	Youth responded to this question on a scale from 1 = "Yes, definitely" to 4 = "No, definitely not." This item was recoded into a binary variable where affirmative responses (definitely, probably) were coded as 1 and negative responses (definitely not, probably not) were coded as 0.
If you have sexual intercourse in the next year, do you intend to use birth control?	Youth responded to this question on a scale from 1 = "Yes, definitely" to 4 = "No, definitely not." This item was recoded into a binary variable where affirmative responses (definitely, probably) were coded as 1 and negative responses (definitely not, probably not) were coded as 0.
If you have sexual intercourse in the next year, do you intend to use a condom?	Youth responded to this question on a scale from 1 = "Yes, definitely" to 4 = "No, definitely not." This item was recoded into a binary variable where affirmative responses (definitely, probably) were coded as 1 and negative responses (definitely not, probably not) were coded as 0.

Skills

The short-term follow-up survey included items regarding skills important to reproductive health. From these, we constructed measures to examine program impacts on youth's perceived ability to say no to sex (*refusal skills*), and successfully negotiate condom use with a partner (*condom negotiation skills*). These measures are described below and detailed information about their component items is presented in Exhibit C2.5.

- ***Refusal skills*** is a composite measure that is the mean of responses to six items about perceived ability to say no to sex in a variety of situations. (See Exhibit C2.5 for coding and other details.) Scores on this scale range from 1 to 4 with high scores indicating more confidence in one's abilities to abstain from intercourse. The measure demonstrates good internal consistency reliability ($\alpha = 0.86$).

- ***Condom negotiation skills*** is a composite measure that is the mean of responses to seven items about perceived ability to obtain and negotiate the use of condoms. (See Exhibit C2.5 for coding and other details.) Scores on this scale range from 1 to 4 with high scores indicating more confidence in one's abilities to obtain and negotiate the use of condoms. The measure demonstrates good internal consistency reliability ($\alpha = 0.83$).

Exhibit C2.5: Skills Scales and Component Items

Component Items	Coding
Refusal Skills (6 items)	
How sure are you that you would be able to say no to having sexual intercourse if your partner really wanted to, but you were not ready?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
How sure are you that you would be able to say no to having sexual intercourse if you just met someone you really liked and that person wanted to have sex, but you didn't?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
How sure are you that you would be able to say no to having sexual intercourse if you had strong sexual feelings for that person?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
How sure are you that you would be able to say no to having sexual intercourse if neither you nor your partner had any form of birth control?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
How sure are you that you would be able to say no to having sexual intercourse if you have dated for a long time?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
How sure are you that you would be able to say no to having sexual intercourse after you have been drinking alcohol?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
Condom Negotiation Skills (7 items)	
If you were going to have sex could you get or buy a condom?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
If you were going to have sex could you talk about using condoms with your partner before having sex?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
If you were going to have sex could you insist on using a condom if your partner didn't want to use one?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
If you were going to have sex could you ask your partner to use condoms even if the two of you had sex before w/o using condoms?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
If you were going to have sex could you use a condom without spoiling the mood?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
If you were going to have sex could you ask a new partner to use condoms?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.
If you were going to have sex could you get a partner to use condoms, even if you're drunk or high?	Youth responded to this question on a scale from 1 = "I'm sure I could" to 4 = "I'm sure I could not." We reverse coded this item so that higher values indicate more confidence in one's ability.

C.3 Youth Sexual Behavior and Sexual Risk

To understand program effects on youths' sexual behavior and sexual risk, we examined their responses to questions about their history of sexual activity, their recent sexual behavior, and their recent sexual risk behavior. We used the seven items presented in Exhibit C.3.1 to examine impacts on sexual behavior and sexual risk.

Exhibit C.3.1: Youth Sexual Behavior and Sexual Risk Measures

Measure	Item	Coding
Sexual Behavior		
Initiation of sexual activity	Have you ever had any of the following: sexual intercourse, oral sex or anal sex?	Youth who were not sexually active at baseline responded to this question with a yes(1)/no(0) answer. This item was coded 0 or 1, with 1 representing one or more forms of sexual activity (sexual intercourse, oral sex, and/or anal sex) during one's lifetime and 0 representing no sexual activity during one's lifetime. Responses to other sexual behavior and sexual risk questions were examined and back-coded into this question such that youth who reported they had engaged in one or more of the sexual activities received a score of 1. Note that sexual activity is defined differently across grantees. In Better Family Life, sexual activity refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in LifeWorks or San Diego Youth Services.
Currently sexually active (in last 90 days)	Coded from three separate items measuring sexual intercourse in the last 90 days, oral sex in the last 90 days, and anal sex in the last 90 days.	Youth who reported they had engaged in one or more of the sexual activities (sexual intercourse, oral sex, or anal sex) during the last 90 days received a score of 1 on this measure. Youth who reported no sexual activity during the last 90 days received a score of 0, as did those who reported (on a separate question) that they had never been sexually active. Note that sexual activity is defined differently across grantees. In Better Family Life, sexual activity refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in LifeWorks or San Diego Youth Services.
Sexual intercourse in the last 90 days	Now please think about the past 3 months. In the past 3 months, have you had sexual intercourse?	Youth responded to this question with a yes(1)/no(0) answer. Youth who reported engaging in sexual intercourse in the last 90 days received a score of 1 on the measure. Those who reported they had not engaged in sexual intercourse in the last 90 days received a score of 0 on the measure, as did those who reported (on a separate question) that they had never been sexually active.
Oral sex in the last 90 days	Now please think about the past 3 months. In the past 3 months, have you had oral sex?	Youth responded to this question with a yes(1)/no(0) answer. Youth who reported engaging in oral sex in the last 90 days received a score of 1 on the measure. Those who reported they had not engaged in oral sex in the last 90 days received a score of 0 on the measure, as did those who reported (on a separate question) that they had never been sexually active.
Sexual Risk		
Sexual intercourse without a condom (in last 90 days)	In the past 3 months, have you had sexual intercourse without you or your partner using a condom?	Youth responded to this question with a yes(1)/no(0) answer. Youth who reported engaging in sexual intercourse without a condom in the last 90 days received a score of 1 on the measure. Those who reported they had not engaged in sexual intercourse without a condom in the last 90 days received a score of 0 on the measure, as did those who reported (on separate questions) that they had not had sexual intercourse in the last 90 days or that they had never been sexually active.

Measure	Item	Coding
Oral sex without a condom (in last 90 days)	In the past 3 months, have you had oral sex without using a condom, even once?	Youth responded to this question with a yes(1)/no(0) answer. Youth who reported engaging in oral sex without a condom in the last 90 days received a score of 1 on the measure. Those who reported they had not engaged in oral sex without a condom in the last 90 days received a score of 0 on the measure, as did those who reported (on separate questions) that they had not had oral sex in the last 90 days or that they had never been sexually active.
Sexual intercourse without birth control (in last 90 days)	In the past 3 months, have you had sexual intercourse without you or your partner using any of these methods of birth control, even just once? • Condoms • Birth control pills • The shot (Depo-Provera) • The patch • The ring (NuvaRing) • IUD (Mirena or Paragard) • Implants (Implanon)	Youth responded to this question with a yes(1)/no(0) answer. Youth who reported engaging in sexual intercourse without birth control in the last 90 days received a score of 1 on the measure. Those who reported they had not engaged in sexual intercourse without birth control in the last 90 days received a score of 0 on the measure, as did those who reported (on separate questions) that they had not had sexual intercourse in the last 90 days or that they had never been sexually active.

Appendix D: Supporting Tables

Exhibit D.1: Characteristics of the Analytic Sample at Baseline

Measure	Treatment Mean[a]	Control Mean	Group Difference[b]	p-value
Demographic characteristics				
Age				
Mean	14.50	14.56	-0.07	0.204
Grade				
Mean	9.25	9.27	-0.02	0.626
Gender				
Female[c]	49.10	49.10	0.00	1.00
Race/ethnicity[d]				
Hispanic	46.09	47.12	-1.04	0.518
Black	33.10	32.96	0.14	0.903
White	11.34	10.73	0.61	0.616
Other	9.54	9.18	0.36	0.792
Family structure and relationships				
Lives with biological parents	93.06	92.24	0.82	0.478
Feels very close to and cared for by father	45.46	46.74	-1.27	0.564
Feels very close to and cared for by mother	63.38	65.98	-2.61	0.171
Risk behaviors				
Ever smoked cigarettes	21.01	20.63	0.38	0.826
Ever drank alcohol	45.73	45.05	0.69	0.743
Ever used marijuana	31.23	30.00	1.23	0.521
Knowledge				
Knowledge of pregnancy risk[e]	51.84	50.61	1.22	0.452
Knowledge of STI risk[e]	44.42	43.46	0.96	0.411
Attitudes				
Attitudes toward protection[f]	3.04	3.04	0.00	0.907
Intentions				
Intentions to have oral sex in the next 12 months	30.00	30.09	-0.08	0.965
Intentions to have sexual intercourse in the next 12 months	41.14	39.16	1.98	0.312
Intentions to use a condom if they were to have sexual intercourse	94.59	94.17	0.42	0.650
Intentions to use birth control if they were to have sexual intercourse	89.41	90.79	-1.38	0.259
Sexual Behavior				
Ever sexually active[g]	30.57	31.32	-0.75	0.683
Currently sexually active (in last 90 days)[g]	18.37	20.79	-2.42	0.175
Sexual intercourse in the last 90 days	16.53	17.99	-1.46	0.409
Oral sex in the last 90 days	12.12	14.56	-2.44	0.114

Measure	Treatment Mean[a]	Control Mean	Group Difference[b]	p-value
Sexual Risk				
Sexual intercourse without a condom in the last 90 days	8.04	9.85	-1.80	0.189
Oral sex without a condom in the last 90 days	10.40	11.77	-1.37	0.348
Sexual intercourse without birth control in the last 90 days	5.36	6.71	-1.35	0.254
Baseline exposure to program information[h]				
Relationships or marriage	81.26	80.02	1.24	0.436
Abstinence from sex	74.31	70.31	4.00*	0.036
Birth control methods	58.78	59.17	-0.40	0.838
Where to obtain birth control	46.26	46.13	0.13	0.956
Sexually transmitted infections	85.06	83.00	2.06	0.157
How to talk with partner about sex and birth control	51.34	52.20	-0.87	0.662
How to say no to sex	71.36	72.91	-1.55	0.379
How babies are made	87.41	88.17	-0.76	0.561

Source: Baseline survey administered prior to randomization.

Notes: Results in this table are based on the analytic sample of 2,368 - 2,689 respondents who provided valid survey responses to relevant items. Values shown are percentages unless otherwise indicated. The items that compose measures of attitudes toward risky behavior, motivation to delay childbearing, refusal skills, and condom negotiation skills were not asked at baseline.

[a] The treatment mean was calculated as the sum of the control group mean and the model estimated treatment-control difference (group difference).

[b] The baseline treatment-control difference was estimated where the dependent variable was the baseline measure, and the only independent variables included in the model were the treatment group indicator and terms for the site/school/semester/gender blocks. Due to rounding, reported group differences may differ from differences between reported means for the treatment and control groups.

[c] The analytic model for outcomes estimates impacts within gender groups, and aggregates impacts across the groups. This approach induces exact baseline equivalence of treatment and control groups on gender.

[d] Racial ethnic categories are Hispanic, black non-Hispanic, white non-Hispanic, and other race non-Hispanic, where other is defined as Asian, American Indian or Alaska native, native Hawaiian or other Pacific Islander, multiracial, or undisclosed race.

[e] Knowledge variables are composite scale scores representing the proportion of items answered correctly.

[f] Attitudes toward protection variable is a composite scale score with higher scores indicating more positive attitudes.

[g] Sexual activity is defined differently across grantees. In one site, sexual activity refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in two sites.

[h] Refers to information received in the 12 months prior to the survey administration.

* p< 0.05, ** p< 0.01, *** p< 0.001 (two-tailed tests).