Reducing the Risk: Final Impact Report

Teen Pregnancy Prevention Replication Study

Contract # HHSP23320095624WC Order # HHSP23337011T

October 2018

Prepared for:
Amy Farb
Office of Adolescent Health
Lisa Trivits
Sarah Oberlander
Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services

Submitted by:

Meredith Kelsey
Jean Layzer
Michelle Blocklin
Cristofer Price

Abt Associates
10 Fawcett Street
Cambridge, MA 02138

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

Recommended Citation

Kelsey, M., Layzer, J., Blocklin, M., and Price, C. "Reducing the Risk Final Impact Report: Findings from the Teen Pregnancy Prevention Replication Study". Cambridge, MA: Abt Associates Inc., October 2018.

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 and Randall Juras provided guidance and support during the initial study design phase, and Beth Boulay provided valuable feedback on the reports. Thanks to Dan Weiss, Heather He, and Jessica Walker for programming and technical assistance, Stefanie Falzone and Bry Pollack for producing and editing 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) is a priority for the U.S. Department of Health and Human Services (HHS). 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 and 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 prevention of teen pregnancy, STIs, and sexual risk behaviors (Kappeler & Farb, 2014).

The TPP Program 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). The review highlighted that the evidence for many of the 28 programs eligible for replication funding rested on single studies 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 TPP Replication Study

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

Impact Study. Through a series of rigorous experimental design evaluations, the impact study tests multiple replications of three evidence-based program models to determine their effectiveness across different settings and populations.

Implementation Study. A comprehensive implementation study provides information about the contexts in which the evidence-based programs were implemented, the challenges faced in implementing them, and aspects of program implementation that help to explain program impacts.

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

1.2 The Three Models Replicated

ASPE and OAH selected three program models from the first round of TPP-funded grants to test and replicate:

- *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 focuses on *Reducing the Risk (RtR),* presenting findings from two follow-up surveys designed to examine the program's short-term and longer-term impacts. It is one in a series of reports that present findings on the implementation and effectiveness of the three program models. Three implementation study reports 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.[3]

[2] Of the 28 program models in the TPP Program, the *Teen Outreach Program (TOP)* was 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] The profiles are available at https://aspe.hhs.gov/basic-report/tpp-replication-study.

2. The Program Model: *Reducing the Risk*

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: (1) initiation of sexual intercourse, (2) abstinence, (3) use of condoms, and (4) use of birth control. *RtR* is intended for use in high school classrooms with students of all ethnicities, although program materials suggest it can also be delivered in community settings. *RtR* consists of 16 units of 45 minutes each. The units 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 birth control 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 unit in which they should be covered (or used, in the case of pedagogical strategies). Exhibit 2.1 shows these curricular components mapped by unit. 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 Unit

Core Content Component	Unit
Knowledge about:	
Pregnancy risk	1
HIV and other STI prevention, transmission, treatment, and consequences	1A,[a] 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
Effective refusal skills and delaying tactics	3, 6
Attitudes about:	
Abstinence	2
Having sex and unprotected sex	3
Using condoms and other birth control	14

Core Content Component	Unit
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/birth control	7
Negotiate 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/birth control	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
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

[a] Implementers of the *RtR* program can choose either Unit 1 or Unit 1A, or if there is time in the schedule, deliver both units to strengthen the message about STI risk delivered in a later unit.

2.2 *Reducing the Risk* Logic Model

RtR is one of 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 is often used in an abbreviated form when schools are unwilling or unable to accommodate its 16 units into their schedules.

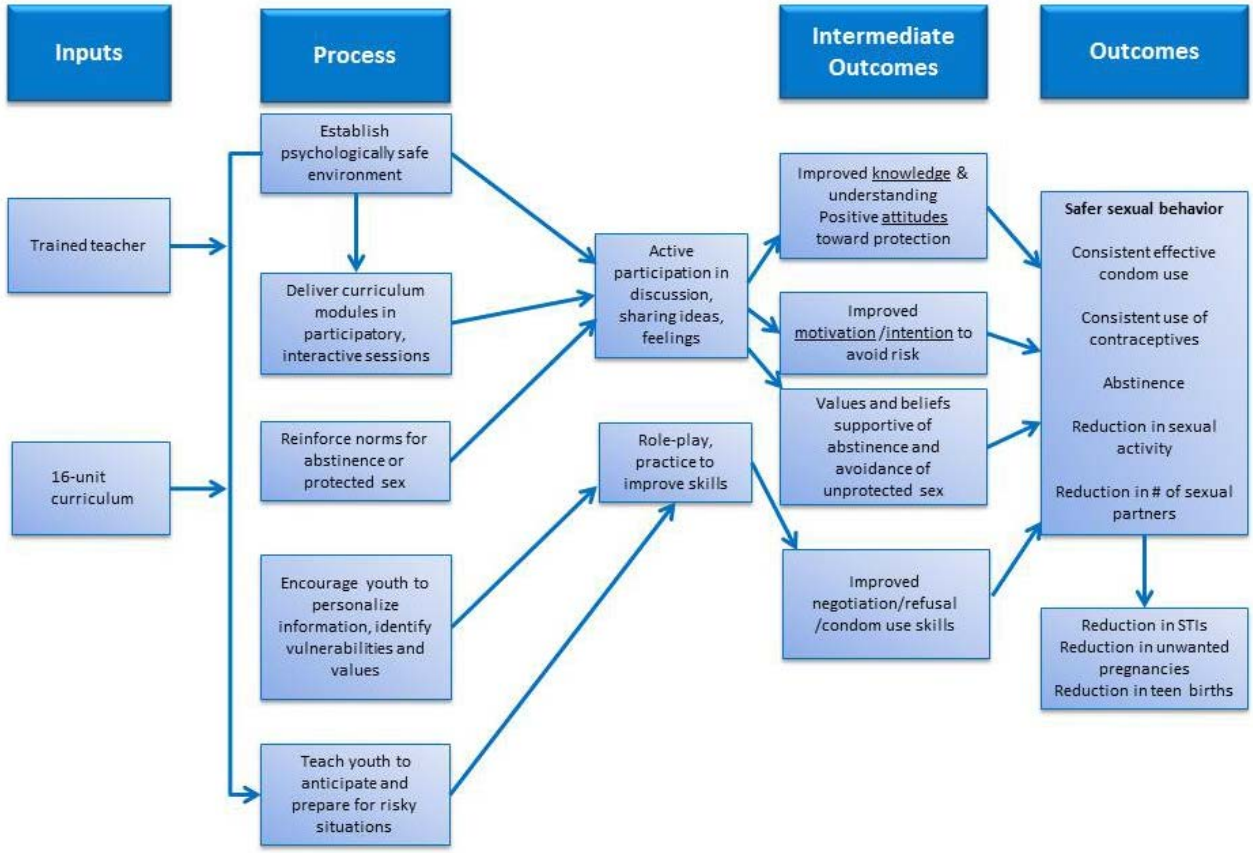
Exhibit 2.2 (below) shows the program elements, the intended outcomes, and the pathways by which the program seeks to achieve these outcomes. The teacher or health educator delivers *RtR* in a classroom or other setting. The first objective for the teacher is to create a psychologically safe environment 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 units in a prescribed sequence. As part of every unit, 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 units 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 units 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, unplanned pregnancies, and births among teens.

Exhibit 2.2: *Reducing the Risk* Logic Model



The model implies but does not specify time periods between early non-behavioral ("intermediate") outcomes and later behavioral outcomes that the early outcomes are believed to influence. It is important to recognize that these intervals may be different depending on the age of program participants and the extent to which they are sexually active. For youth at the lower end of the age range for which the program is suitable (e.g., age 13-14), *RtR*, like many other teen pregnancy prevention programs, may operate as a true ***prevention*** program. That is, its messages are conveyed to younger youth for the most part before they become sexually active in order to prevent unsafe sexual behavior and its consequences. The model posits that youth acquire knowledge and develop positive attitudes, motivation, intentions, and skills to avoid or prevent risk behavior first. Then the opportunity to translate that knowledge, intentions, and skills into action may not arise for *several years*, when youth are more likely to become sexually active.

By contrast, if the program is implemented with youth at the higher end of the age spectrum (e.g., age 17-19), it could act as an ***intervention*** for a majority of the participants, those who are sexually active. That is, its aim would be to change behaviors that youth are already engaged in, so those behaviors are less risky, over a much shorter period of time. If the program is able to affect positive changes in knowledge, attitudes, motivation, intentions, and skills among older, sexually experienced youth, then these non-behavioral changes should translate *more quickly* into changes in sexual risk behavior and the consequences of that behavior.

Though the general guidance is that *RtR* is appropriate for a wide age range, it is important to recognize that the ability to detect behavioral changes for the entire participant group, across ages, may be limited. The wide range in age of participants and the variation in baseline levels of sexual experience may obscure potentially different prevention and intervention effects. We will revisit this issue later in the report as we present and discuss the study's findings.

In Section 3.5.2, we describe in detail the modifications to the *RtR* program model proposed by each of the organizations replicating it. They made modifications to comply with district policies, to address gaps in program content, or to accommodate school concerns or local constraints, though still adhering to the core elements of the model.

2.3 Prior Evidence of Effectiveness

RtR is one of the programs identified in the research as having evidence of effectiveness, and therefore was eligible for replication funding under the TPP Program (HHS, 2010). The evidence for *RtR*'s effectiveness making it eligible for replication came from one study that was completed many years ago (25 years ago, in this case).

That previous study of *RtR* by Kirby and colleagues (1991) employed a quasi-experimental design and was conducted in 46 classes in rural and urban schools in northern California. 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 ninth graders, and the remainder were 11th and 12th graders. Students were surveyed six and 18 months after the intervention ended. Analysis of the survey data found the following:

- *RtR* had no impacts on recent[4] sexual activity at six months. However, at 18 months, among students who were sexually inexperienced at baseline, significantly fewer *RtR* participants had initiated sexual intercourse than students in the comparison group.
- *RtR* had no effects on unprotected sex at six 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; this effect was driven primarily by females and lower-risk *RtR* participants.
- *RtR* had significant positive effects on participants' knowledge of pregnancy and STI risk and on their perceptions of the proportion of their peers who had ever had sexual intercourse (Kirby, Barch, Leland, & Fetro, 1991).

There have been additional evaluations of *RtR* since the initial evidence review that have evaluated the program as originally designed, as well as adaptations. Some of these evaluations found no evidence of program effectiveness, whereas others revealed some favorable impacts, notably on sexual initiation (Barbee, Cunningham, van Zyl, Antle, & Langley, 2016; Reyna & Mills, 2014).

[4] *Recent* was defined as in the last 90 days.

3. Evaluation Design

The impact study is designed to estimate the effects of *RtR* on sexual risk behaviors and consequences, as well as on the non-behavioral, intermediate outcomes the logic model predicts will lead to the behavioral outcomes that *RtR* seeks to achieve.[5]

In the first part of this chapter, we set forth the study's research questions and describe the design elements of the study, including the overall evaluation strategy; the measures selected to address the research questions and the timing of measurements; and the analytic strategy devised to assess program effectiveness. In the second part of the chapter, we describe our implementation of the study design and analysis plan in the replication sites.

3.1 Research Questions

The evaluation is guided by the following research questions.

1. Did *RtR* have an impact on sexual behavior after 12 months and 24 months?
2. Did *RtR* reduce the incidence of unplanned teen pregnancies after 24 months?
3. Did *RtR* reduce the incidence of STIs after 24 months?
4. Did *RtR* have an effect on non-behavioral, intermediate outcomes hypothesized to lead to behavior change (i.e., knowledge, attitudes, motivation, intentions, and skills) after 12 months and 24 months?
5. Do program effects on behavior differ by replication site and for key subgroups (e.g., gender, age, race/ethnicity, sexual experience at baseline)?

These five research questions imply a wide range of outcomes, including non-behavioral (intermediate) outcomes that the program model suggests are precursors of the behavioral outcomes, and the behavioral consequences that are the ultimate targets of the program and the TPP Initiative. The fifth research question is intended to take maximum advantage of pooled data from all three replications by exploring potential differences in effect for specific sites and subgroups.

We elected to investigate non-behavioral as well as behavioral outcomes for two reasons: first, because we hoped to be able to trace the pathways of change set forth in the program logic model; and second, because many youth participating in the study, particularly those who are younger, may not become sexually active even at the study's final measurement point. Effects on non-behavioral outcomes, as we explained earlier, are the main outcomes we can observe for these young people.[6]

Collecting data and estimating effects on so many outcomes does, however, pose challenges for the ways in which data are analyzed and how results are interpreted. The sheer number of statistical tests of effectiveness means that we would expect five percent to generate statistically significant results simply

---

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

[6] Note that this is true only for those who were sexually inexperienced when they entered the study and remained so throughout the study period.

by chance. In a later section of this chapter we describe the steps we took to minimize the risk of incorrectly concluding that *RtR* had an impact.

3.2 Key Design Features

The design of the evaluation of *RtR* included the following key elements:

- Multiple replications of the program model (three sites).
- Within each replication site, implementation of a rigorous experimental design in which classes were randomly assigned either to receive the *RtR* intervention or to a control group that received the usual curriculum (e.g., physical education, science, health, or other non-core subject).
- Construction of measures that allow us to address all of the research questions.
- A measurement schedule that captures both short-term and longer-term outcomes.
- An analytic strategy that pools data from all replications to allow us to measure sexual behavior and the consequences of sexual risk behavior and to examine differences in program effectiveness by replication site, as well as for important youth subgroups.
- A strategy that identifies a set of key behavioral outcomes and prioritizes a limited number of confirmatory analyses to increase confidence in the study findings. At the same time, the strategy also allows for exploratory (and more speculative) analyses that incorporate many more outcomes, both behavioral and non-behavioral.

3.3 Measures and Measurement Schedule

Outcome measures selected for the study fall into three major categories: behavior, consequences, and non-behavioral intermediate outcomes. Exhibit 3.1 summarizes the outcome measures and their construction; a more complete description of each measure and its individual items can be found in Appendix A.

Exhibit 3.1: Outcome Measures

Measure	Definition
Sexual Behavior Outcomes	
Sexual activity	
• Recent sexual activity (in last 90 days)[a] • Sexual intercourse in the last 90 days • Oral sex in the last 90 days	Single items, scored 1 (yes) or 0 (no).
• 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, 1 (yes) or 0 (no).
Sexual risk behavior	
• 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, scored 1 (yes) or 0 (no).
Sexual Consequences (longer-term follow-up only)	
• Pregnant or gotten someone pregnant since baseline[a] • Diagnosed with STI in the last 12 months	Single items, scored 1 (yes) or 0 (no).

Measure	Definition
Non-Behavioral Intermediate Outcomes	
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, multiplied by 100. Average scores 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. Average scores range from 0 to 100 and represent the percentage of the 12 questions answered correctly, with higher values representing more accurate knowledge.
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 sexual behavior.
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.
Intentions (in next 12 months)	
• Intention to have sexual intercourse • Intention to have oral sex • Intention to use birth control if having sexual intercourse • Intention to use a condom if having sexual intercourse	Single items, scored 0 or 1, with 1 representing stronger intention.
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.
• 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.

[a] Designated as a key outcome for confirmatory analyses (see Section 3.4.2).

The study design called for youth in the three replication sites to be surveyed three times: before the intervention began (baseline); 12 months after the baseline survey (short-term follow-up); and 24 months after the baseline survey (longer-term follow-up). This schedule allowed us to capture short-term outcomes that might not persist after a longer interval. It also let us capture, in the longer term, behavioral outcomes that may take longer to emerge, particularly for youth who were not yet sexually active at the time of the short-term follow-up.

3.4 Analytic Approach

Two strategic decisions shaped the analysis of data collected over the life of the study. The first was a decision about how to treat the three replications of the program. The second was a decision about

prioritizing analyses to answer the key research questions. Each of these decisions as they relate to our analytic approach is described below.

3.4.1 Incorporating Three Program Replications

When deciding how to treat the three replications of *RtR*, one possibility was to treat them as ***three stand-alone evaluations***. Abt staff designed each of the three evaluations independently, taking into account any special circumstances in each replication site (e.g., in two of the replication sites, at those grantees' request, surveys of youth excluded questions about anal sex). The sample requirements in each of the replication sites were calculated to permit detection of relatively small impacts on sexual behavior.

The other possibility—the one ultimately selected—was to consider the three evaluations as ***components of one integrated study***, in which data were pooled across the three sites. This strategy offered several benefits. Importantly, the tripled sample size would allow us to estimate the impact of *RtR* on likely consequences of sexual risk behavior such as pregnancy and STIs. Prevention of these consequences is the primary goal of the TPP Initiative, but measuring them as part of an evaluation is a challenge. Given that these outcomes are relatively rare events, the sample size necessary to detect a possible intervention impact on pregnancies or STIs requires resources beyond what is available in many single-site studies.

In addition, pooling data across the three replication sites would allow us to conduct the many subgroup analyses necessary to address the study's research questions. Subgroup analyses would be less feasible with the smaller sample sizes of the individual replications. Even with pooling the data across sites, we also have the ability to examine the extent to which replications differed in their effectiveness.

Finally, although three replications cannot be held to represent the universe of possible replications, findings from the analysis of pooled data would have greater generalizability than findings from any single-replication study. An integrated study would include a variety of settings, a range of student ages, and variation in other demographic characteristics.

A decision to create an integrated evaluation in which data from all three replications would be pooled for analytic purposes was supported by OAH's requirements of grantees to define, measure, and adhere to fidelity to the program model. These requirements ensured that each of the three replications implemented the same core program elements. The random assignment, measurement, and data collection procedures were also the same across the replication sites. The consistency of these design elements ensured that impact estimates derived from data pooled at the program level would represent rigorous tests of a well-defined and well-implemented program model.

For all these reasons, we elected to pool the data from all three replication sites.

3.4.2 Prioritizing the Analyses Needed to Answer Key Research Questions

We noted earlier that the study's research questions demonstrate interest in a variety of outcomes, both behavioral and non-behavioral, as well as interest in understanding the extent to which the program works differently for different replication sites and different subgroups. In practical terms, exploring these multiple interests translates into a large number of statistical tests, of which some predictable share will produce statistically significant impact findings simply by chance. To reduce the risk of spurious findings, we needed to develop a strategy that assigned the greatest weight to analyses of greatest interest to federal policymakers.

The first step was to identify a small set of behavioral outcomes by which the success of *RtR* would be judged. These outcomes reflect the goals of the federal TPP Program and of most of the interventions funded by it. These outcomes span both short- and longer-term measurement points. Exhibit 3.2 shows the measurement domains and the key outcomes we identified.

Exhibit 3.2: Measurement Domains and Key Outcomes

Measurement Domain	Outcomes
Recent sexual behavior at the short-term follow-up	1. Sexual activity in the last 90 days 2. Sexual intercourse without birth control in the last 90 days
Recent sexual behavior at the longer-term follow-up	1. Sexual activity in the last 90 days 2. Sexual intercourse without birth control in the last 90 days
Consequences of sexual risk behavior	1. Pregnancy since baseline[a]

[a] The pregnancy outcome was reported only at the longer-term follow-up because of the low prevalence rate and statistical considerations.

The second step was to identify the sample on which to test impacts of *RtR* on these key outcomes. Given the advantages of a large, diverse sample, we selected the full sample, pooling data across the three replication sites.

The analyses used to make claims about the impact of *RtR* are referred to as "confirmatory analyses." Based on the decisions described above, our confirmatory analyses estimate the impacts of *RtR* on the *key outcomes* for the *full sample*, using data pooled across the three replication sites. Additional analyses, testing different outcomes or using different samples or subgroups, are referred to as "exploratory analyses" and should be interpreted as suggestive of potential program effects (see Schochet, 2008a).

In Section 3.6, we describe in more detail how the impact analyses were conducted and the procedures for making statistical corrections for conducting multiple tests.

3.5 Implementing the Study Design

This section describes the selection of the three replication sites, site-specific program designs, settings for the program, the treatment and control conditions, recruitment and random assignment, and our data collection strategy.

3.5.1 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 that most grantees had not planned for a rigorous evaluation.[7] 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

[7] 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, a larger-scale replication, had proposed a rigorous local evaluation.

honor the agreements if that meant participating in a rigorous evaluation. In still other cases, grantees were struggling to reach agreements with school districts just 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 are described below.

- **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** 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 (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 across the county.

3.5.2 Site-Specific Program Designs

The three replications of *RtR* shared important aspects of the program model. All three replications were required to implement the program with fidelity to the core elements of the model (shown in Exhibit 2.1). Fidelity was assessed, monitored, and reported to OAH at regular intervals by program staff. Grantees also proposed to deliver all 16 units of the *RtR* program. However, there were small variations in program design across the three sites.

In its grant proposal, BFL requested and received permission from OAH for one school to deliver the program to male and female students separately, each with a health educator of that 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 units: one on reproductive anatomy and the optional *RtR* introductory unit, reinforcing messages about pregnancy and STI prevention. LifeWorks was also permitted to replace *RtR*'s condom demonstration with printed materials, to conform to district policy. SDYS received permission to deliver the program with two health educators, to accommodate larger class sizes.

3.5.3 Settings for the Program

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 the city of St. Louis, St. Louis County in Missouri, and in St. Clair County in 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 (AISD). SDYS and its four partners implemented the program in 8th or 9th grade physical education, health, or science classes in six public middle, junior high, and high schools in San Diego County, California.

3.5.4 Treatment and Control Conditions

Across the three replications, students assigned to the treatment groups were offered all 16 units[8] of *RtR*, delivered by health educators[9] hired by the grantee or a partner agency. Health educators were trained by the program distributor and then supervised and monitored continuously by supervisory staff in each agency and local external evaluators. Fidelity, quality, and attendance measures required by OAH were completed and reported by the health educators, supervisory staff, and evaluators.

Beyond these commonalities, there were minor differences in the treatment and control conditions across replication sites, as shown in Exhibit 3.3. The treatment condition varied in number of units delivered, schedules for delivering the curriculum, and class size. The control condition represented business-as-usual in each of the replication sites: In one of the replication sites, control group students 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 3.3: Treatment and Control Conditions in the Three Replications

Grantee	Treatment Group	Control Group
Better Family Life	Number of units delivered: 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)	Physical education, ROTC, health, homeroom/guidance classes at schools' discretion
LifeWorks	Number of units delivered: 18 Delivery schedule: Delivered in nine 90-minute sessions Gender and size of groups: Average-size, mixed-gender classes	Health class
San Diego Youth Services	Number of units delivered: 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	Physical education, health, or science classes at schools' discretion

[8] LifeWorks staff also delivered the additional two sessions described earlier.

[9] Depending on the site, the staff who delivered *RtR* were variously called "health educators" or "facilitators." We have used *health educators* as a blanket term in this report.

3.5.5 Recruitment and Random Assignment

In each replication site, grantees recruited schools to participate in the study. 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).

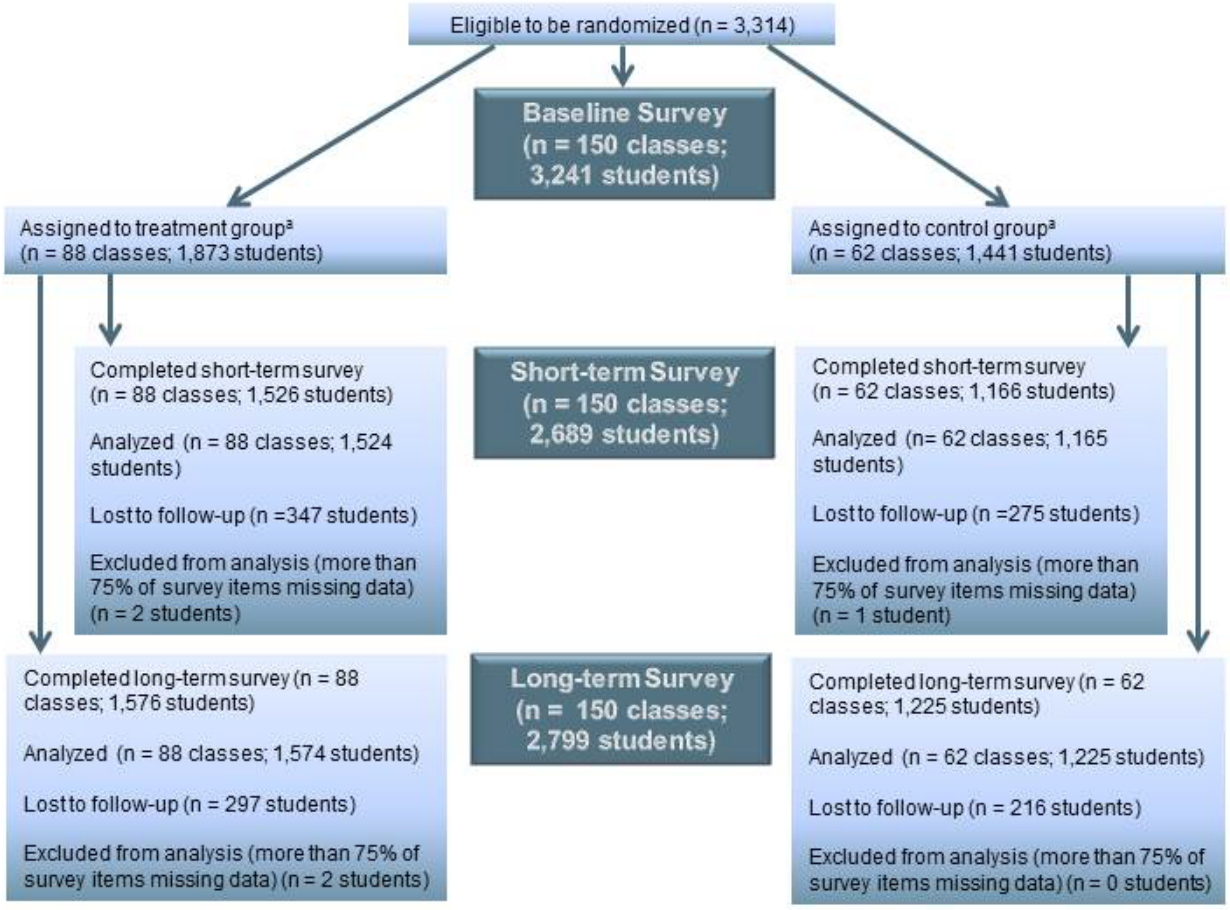
Grantee and partner staff who had been carefully trained by Abt study staff conducted presentations to each class. These presentations included information about the study procedures and a description of the treatment and control conditions. The presentations were intended to personalize the study and 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. Grantee and partner staff worked with individual teachers to obtain parental consents, and notified the Abt team about students whose parents had consented to their child participating in the study (i.e., "eligible" students). Students whose parents had not given consent were reassigned to a different class scheduled for the same class period.

Once the period allowed for consent had expired for a given school, a baseline survey was administered to all eligible students.

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 grantee or partner preferences for program delivery, 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. Teachers, schools, and students were informed of the random assignment results only after students had completed the baseline survey.

Exhibit 3.4 shows how we arrived at the study's analytic samples (i.e., short- and longer-term analytic samples) 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).

Exhibit 3.4: Study Sample



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

3.5.6 Data Collection Strategy

A web-based Audio Computer-Assisted Self-Interview (ACASI) system was used to capture and store survey responses, and youth 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.

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

For the two follow-up surveys, 12 months after baseline (short-term) and 24 months after baseline (longer-term), only the web-based ACASI system was used. Study participants could access and complete the survey using personal tablets or computers, school or library computers, or even their smart phones. We sent text reminders to all study participants (regardless of whether they had completed the baseline survey) before a survey went live and then throughout its three-month survey period. For study

participants who had not yet completed the survey near the end of the survey period, field staff contacted them and encouraged them to complete the survey independently online or helped them to access the survey. Gift cards were mailed to participants after survey completion of the survey.

As Exhibit 3.5 shows, 81 percent of eligible students completed the short-term follow-up survey and 85 percent completed the longer-term follow-up survey. At both data collection points, there was almost no difference in the response rates of students in the treatment groups versus those in the control groups. Of the three replication sites, BFL had the highest response rate at the short-term follow-up, and SDYS had the highest response rate at the longer-term follow-up.

Exhibit 3.5 *Reducing the Risk* Survey Response Rates

	All Participants			Completed Short-Term Follow-Up						Completed Longer-Term Follow-Up					
				Total		Treatment		Control		Total		Treatment		Control	
	Total N	T	C	N	%	N	%	N	%	N	%	N	%	N	%
All sites	3,314	1,873	1,441	2,689	81.1	1,524	81.4	1,165	80.9	2,799	84.5	1,574	84.0	1,225	85.0
Better Family Life	1,050	640	410	941	89.6	572	89.4	369	90.0	857	81.6	527	82.3	330	80.5
LifeWorks	1,093	568	525	853	78.0	442	77.8	411	78.3	901	82.4	459	80.8	442	84.2
San Diego Youth Services	1,171	665	506	895	76.4	510	76.7	385	76.1	1,041	88.9	588	88.4	453	89.5

3.6 Conducting the Analysis

In this section, we describe in greater detail the analytic procedures used to address the research questions.

3.6.1 Estimation of Impacts for the Full Sample

We estimated program impacts by comparing the outcomes of treatment and control group members using a regression framework, in which we included 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, Olsen, Bell, & Price, 2009). For each outcome measure, the model produces an estimate of the average treatment impact of *RtR* across the three replication sites.

Because random assignment occurred at the classroom level within randomization blocks, we estimated the impacts of *RtR* using a two-level fixed effects model that includes a series of indicator variables representing each of the randomization blocks defined by site, school, semester, and gender. The two-level model (students nested within classrooms) accounts for the clustering of students within classrooms and increases the standard errors of the impact estimates.

The model has the basic structure of equations 1-4 below. In this model, individual outcomes are modeled at Level 1, whereas 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

randomization blocks.[10] The Level 2 model includes a treatment indicator and random intercept terms for classes to account for correlation of individuals within classes.

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

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

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

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

In this model:[11]

Y_{ij} is the outcome of interest (e.g., sexual intercourse without birth control) for the *i*th student in the *j*th class, *m*th randomization 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, Other), risk behaviors (smoking, alcohol use, marijuana use), baseline sexual activity (ever sexually active), baseline knowledge (pregnancy and STI risk), baseline intentions to have oral sex and sexual intercourse, and the baseline measure of the outcome when available;

D_{mij} is a dummy variable representing the *m*th randomization block. These block indicators reflect that there were different treatment probabilities across blocks. Because random assignment blocks were constructed based on site, school, semester, and gender, the dummy variable also accounts for these factors; and

ε_{ijs} and μ_{0js} are random error terms.

The coefficient γ_{01} represents the average impact of the program on the outcome. The *p*-values reported for impact estimates are two-tailed to account for the possibility that the intervention might adversely affect one or more of the outcomes. Criteria for statistical significance and procedures for accounting for multiple hypothesis testing are described in the section that follows. The coefficients on the covariates, β_{kj}, reflect the relationship between the outcome measure and each of the covariates while controlling for others. It is important to note that this model specification treats randomization blocks (and thus sites) and

[10] The block dummies appear in the Level 1 model because there are male and female blocks within classes. Information about sites is also contained within the block dummies. There are no specific model terms for sites because the block dummies are linear combinations of the site indicators.

[11] 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, Clark, Tuttle, & Dwyer, 2010). There were no substantive differences in the inferences that result from any of the three modeling approaches.

the treatment effects as fixed as opposed to random, which is consistent with how the replication sites were chosen and how the results of the study are interpreted.[12]

Equations 1-4 estimate the impact of access to *RtR*. The crucial difference between the treatment and control groups is that they were randomly assigned to receive access to *RtR* services or not. 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 average impact on treatment group members who had the opportunity to participate in the intervention, not the average impact on program group members who actually participated in the intervention. However, most treatment group members received at least 75 percent of the program.

3.6.2 Correcting for the Large Number of Comparisons Needed to Answer Key Questions

As mentioned previously, the confirmatory analyses estimate impacts on the key outcomes for the full sample, using data pooled across the three replication sites. Prioritizing these analyses limits the number of hypothesis tests we conduct to draw causal conclusions, thereby mitigating the risk of incorrectly concluding that *RtR* was effective.

Typically, we use a *p*-value criterion of .05 to determine whether an impact estimate is statistically significant and unlikely to be a chance finding. However, for confirmatory analyses of the outcome domains *recent sexual behavior at the short-term follow-up* and *recent sexual behavior at the longer-term follow-up*, each of which had multiple outcomes, we also applied corrections for multiple comparisons. Specifically, we applied a correction described by Benjamini & Hochberg (1995) that adjusts the criterion used for determining statistical significance to account for multiple tests. Within each of these two domains, the correction means that both of the tests would be deemed significant if both had *p*-values below .05. If only one had a *p*-value below .05, it would be deemed significant only if its *p*-value was below .025. For the *consequences of sexual risk behavior* outcome domain, there was only one key outcome measure, so no multiple comparisons correction was applied. In this domain, we applied the traditional criterion for statistical significance of $p<.05$.

For exploratory analyses (i.e., all non-confirmatory analyses), no adjustments were applied to the criterion for statistical significance.[13] For each exploratory test, we applied the traditional criterion of $p<0.05$ for statistical significance. As noted previously, exploratory analyses should not be used to make causal conclusions about the effectiveness of *RtR*. The results from exploratory analyses are reported separately

---

[12] 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, 2008b, p. 70).

[13] The decision not to apply an adjustment for multiple comparisons to the results of the exploratory analyses aligns with standards of good practice (see Schochet, 2008a) and was made after weighing the risks and benefits. The risk of not applying adjustments for multiple comparisons in the exploratory analyses is the likelihood of spurious findings, and we warn readers about this repeatedly. Conversely, if we were to apply multiple comparisons adjustments to the exploratory findings, the adjustments would be very conservative and practically no results would be flagged as significant. The benefit of reporting unadjusted results from the exploratory analyses is that unadjusted test results help us to identify potentially important findings that may, in turn, help us to interpret the findings from the confirmatory analyses. Unadjusted test results also may suggest promising avenues for future research.

from the results of confirmatory analyses, and readers should interpret exploratory results with caution, keeping in mind that with a large number of tests conducted, the likelihood of obtaining some statistically significant results by chance is high. Even if the intervention had no true impact, we would expect 5 percent of the tests to be significant by chance alone.

3.6.3 Site-Level Analyses

For one set of exploratory analyses, we estimated effects for each site separately and tested for differences in effects between the three sites by including treatment-by-site interaction terms in Equation 1 above (see Section 3.6.1) and testing for the joint significance of the interaction terms. When statistically significant differences are found between sites for one or more outcomes, we discuss these differences. The purpose of testing for differences between sites before discussing site-level results in the main text is to guard against overinterpretation of findings, some of which would be expected by chance in such a large group of outcomes. We discuss site-specific effects only when differences in effects between sites are found, because it is only credible to report an effect in one site—but not in another—if there is a significant difference between the sites. The site-specific results in Appendix B are not adjusted for multiple comparisons, and any significant findings reported there should therefore be interpreted with caution.

3.6.4 Subgroup Analyses

In addition to the overall and site-level effects, we estimated effects for key subgroups of participants—based on gender (male/female), age (less than age 15 / age 15 or older), race/ethnicity (Hispanic, Black, White, Other), and baseline sexual experience (never sexually active at baseline / ever sexually active at baseline[14])—and tested for differences between subgroups, to better understand what works for whom. We implemented subgroup analyses by including subgroup indicators and treatment-by-subgroup interaction terms in Level 1 of the model (i.e., Equation 1 above in Section 3.6.1) and testing for significance of the interaction term.

To guard against potential overinterpretation of results among the large number of subgroup estimates, we present impact estimates for individual subgroups in Appendix C only when there is a statistically significant difference between subgroups; for example, the impact estimate would be presented for the subgroup of boys only if there were a statistically significant difference in impacts between boys and girls.

3.6.5 Approach to Handling Missing Data

We used 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).

[14] Preliminary descriptive analyses to assess differences by subgroups based on youth sexual activity throughout the study are described in Appendix D.

4. Implementation Findings and Baseline Characteristics

Before presenting the impact results of the study, we first consider some important contextual factors that might affect the interpretation of the findings. Both how well a program model is implemented and the characteristics of the population served can strongly influence the extent to which the program is able to meet its goals.

Key Findings

Across the three replication sites,

- *RtR* was implemented as intended, and participants received a majority of the intervention.
- Students varied in their demographic characteristics, engagement in risk behaviors, knowledge and intentions, and sexual behavior at baseline.

Implementation of *RtR* was guided by fidelity requirements established by OAH at the outset of the grant award. The guidelines allow for an assessment of the extent to which the program was implemented with fidelity and to highlight areas where there were differences in implementation across replication sites. In this chapter, we discuss program implementation and then describe the study sample at baseline (i.e., when students were enrolled in the study).

4.1 Program Implementation

As we noted in Section 1.3, a separate report provides 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. In this section, we provide a summary of implementation findings that are directly relevant to the impact findings reported in the next chapters.

RtR 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 high.

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 by 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 unit 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 a degree of uniformity across sites replicating the same program model.

The approved minor modifications described in Section 3.5.2 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. In all three replications, a majority of students received at least three-quarters of the sessions offered. SDYS reported that 85 percent of students attended at least 75 percent of classes; LifeWorks reported 81 percent of students attended at least 75 percent of classes; and BFL reported 73 percent of students attended at least 75 percent of classes.

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 Sample Characteristics

In this section, we present the baseline characteristics of the impact study analytic samples pooled across all three sites, as well as for each individual site. We then describe the comparability of the treatment and control groups at baseline.

4.2.1 Analytic Samples

Baseline characteristics of the longer-term analytic sample for *RtR* overall and for each replication site are presented in Exhibit 4.1.[15]

Age. 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.

Race/Ethnicity. Overall, almost one-third of study participants were non-Hispanic Black; slightly less than half were Hispanic, and the remainder were non-Hispanic White or Other. However, there were large differences

[15] The baseline characteristics of the short-term analytic sample differ little if at all from those shown in Exhibit 4.1. For interested readers, the baseline characteristics of the treatment and control groups in both the short-term analytic sample and the longer-term analytic sample are shown in Appendix Exhibits E.1 and E.2.

between the three replication sites in the racial/ethnic composition of the study samples. The BFL sample was almost entirely Black; whereas Black students made up less than 10 percent of the sample in the two other sites. By contrast, Hispanic students made up about two-thirds of the samples in SDYS and LifeWorks, compared with less than 3 percent in BFL.

Family Structure. 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.

Risk Behaviors. 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 SDYS, where the sample was younger, significantly fewer students had engaged in any of these risk behaviors.

Sexual Activity/Risk Behavior/Consequences. 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. Though less than a third of the overall sample had ever been sexually active, and one-fifth were recently sexually active (i.e., sexually active in the last 90 days), the SDYS sample was strikingly less sexually experienced: Just 12 percent had ever been sexually active ,and less than 10 percent were recently sexually active. Students in BFL consistently reported the highest levels of sexual activity. This same pattern was repeated for sexual risk behaviors and for consequences. The proportions of students who had engaged in unprotected sexual activity, who had been (or gotten someone) pregnant, or had been diagnosed with an STI were consistently lowest in the SDYS sample and highest in the BFL sample.

Knowledge/Attitudes/Intentions. Slightly more than half of the students in the overall sample at baseline demonstrated an accurate understanding of pregnancy risk, whereas a somewhat smaller proportion (44 percent) understood STI risks. On both topics, the LifeWorks sample was more informed than were 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 students 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.

Exhibit 4.1: Baseline Characteristics of the Longer-Term Analytic Sample 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	14.71	15.18	13.72	14.49	.000	***
Grade	9.19	9.83	8.77	9.24	.000	***
Gender						
Female	48.89	49.39	52.26	50.30	.276	
Race/Ethnicity[b]						
Hispanic	2.68	63.15	68.97	46.80	.000	***
Black	89.61	9.43	5.48	32.51	.000	***
White	0.82	22.09	10.18	11.15	.000	***
Other	6.88	5.33	15.37	9.54	.000	***

Measure	Better Family Life	LifeWorks	San Diego Youth Services	*RtR* Overall	*p*-Value for the Test of Differences Across Sites[a]	
Family structure and relationships						
Lives with biological parent/s	92.16	93.12	92.58	92.63	.745	
Feels very close to and cared for by father	45.76	43.57	49.22	46.29	.062	
Feels very close to and cared for by mother	67.43	61.92	64.92	64.71	.057	
Risk behaviors						
Ever smoked cigarettes	18.93	31.21	13.78	21.08	.000	***
Ever drank alcohol	49.11	57.61	33.57	46.23	.000	***
Ever used marijuana	33.41	41.26	18.53	30.56	.000	***
Sexual activity						
Ever sexually active [c]	47.59	38.46	12.00	31.50	.000	***
Recently sexually active (in the last 90 days) [c]	31.28	23.19	7.56	19.93	.000	***
Sexual intercourse in the last 90 days	27.66	20.70	6.65	17.68	.000	***
Oral sex in the last 90 days	19.30	16.33	5.65	13.32	.000	***
Sexual risk behavior						
Sexual intercourse without birth control in the last 90 days	8.94	7.69	2.02	5.99	.000	***
Sexual intercourse without a condom in the last 90 days	12.92	12.56	2.52	8.99	.000	***
Oral sex without a condom in the last 90 days	14.48	14.40	4.74	10.88	.000	***
Consequences of sexual risk behavior						
Ever pregnant or gotten someone pregnant (lifetime)	5.44	3.28	1.31	3.22	.000	***
Diagnosed with STI in the last 12 months	1.09	0.45	0.20	0.55	.036	*
Knowledge, attitudes, and intentions						
Knowledge of pregnancy risk[d]	46.28	60.18	48.71	51.69	.000	***
Knowledge of STI risk[d]	44.52	49.39	38.43	43.87	.000	***
Attitudes toward protection[e]	3.04	3.04	3.05	3.04	.605	
Intentions to have sexual intercourse[f]	52.11	46.99	22.72	39.61	.000	***
Intentions to have oral sex[f]	33.57	37.81	20.14	30.00	.000	***
Intentions to use birth control if they were to have sexual intercourse[f]	87.94	88.96	92.86	90.05	.001	**
Intentions to use a condom[f]	94.41	93.91	94.58	94.30	.813	

Source: Baseline survey administered prior to randomization.

Notes: Data in this table are based on 2,452–2,799 longer-term survey respondents who provided valid survey responses to relevant items on the baseline survey. Baseline characteristics of short-term survey respondents were similar. Values shown are percentages unless otherwise indicated. The items that compose measures of attitudes toward risky sexual 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 race 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/or anal sex. In LifeWorks and San Diego Youth Services, students were not asked about anal sex.

[d] Scores represent the average percentage of items answered correctly.

[e] Scale score averages responses ranging from 1 to 4. Higher scores indicate more positive attitudes.

[f] Intention to engage in the behavior in the next 12 months. Dichotomous variables, reported as percentage of respondents who responded affirmatively.

* $p<.05$; ** $p<.01$; *** $p<.001$ (two-tailed tests).

4.2.2 Comparability of the Treatment and Control 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 for both the short-term and longer-term analytic samples using a series of models with the same structural components as the impact model in Equation 1 in Chapter 3, Section 3.6.1 (i.e., the same randomization block indicators and treatment group indicator), but where, in each model, one of the baseline characteristics 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 baseline measure. None of these differences was significant for either analytic sample (see Appendix Exhibits E.1 and E.2).

5. Program Impacts on Youth Sexual Activity, Sexual Risk Behavior, and Consequences of Sexual Risk Behavior

In this chapter, we present findings from both the short-term and longer-term follow-up surveys on behavioral outcomes. The findings presented here reflect our analytic strategy of first conducting confirmatory analyses on a key set of outcomes to produce findings that are conclusive about the impacts of *RtR* rather than suggestive.

We conducted additional analyses to explore program effects on other, related sexual behaviors, sexual risk behaviors, and consequences. We begin this chapter with a discussion of the confirmatory analyses, followed by a presentation of program effects on other behaviors for the full sample. Findings for site-level impacts and specific subgroups of interest are discussed in Chapter 7.

Key Behavioral Impact Findings

Confirmatory analyses revealed no significant impacts of *RtR* on:

- Recent sexual activity; and sexual intercourse without birth control in the short-term (12 months after baseline).
- Recent sexual activity; and sexual intercourse without birth control in the longer-term (24 months after baseline).
- Pregnancy (between the baseline and 24-month follow-up surveys).

Exploratory analyses revealed no significant overall effects of *RtR* on:

- Recent sexual intercourse; recent oral sex.
- Initiation of sexual activity.
- Recent sexual intercourse without a condom; recent oral sex without a condom.
- Recent diagnosis of a sexually transmitted infection.

5.1 Confirmatory Analyses of Program Impacts on Key Behavioral Outcomes

The pre-specified confirmatory analyses tested the impacts of *RtR* on the following key outcomes for the full sample: recent sexual activity and sexual intercourse without birth control in the short term (12 months after baseline); recent sexual activity and sexual intercourse without birth control in the longer term (24 months after baseline); and pregnancy (between the baseline and 24-month follow-up surveys). In order to minimize the concern that our confirmatory analyses would miss a behavioral impact that occurred earlier in the follow-up period but nonetheless affected pregnancy, we treat recent sexual behavior at the short-term follow-up as distinct from recent sexual behavior at the longer-term follow-up (see Exhibit 5.1).

Confirmatory analyses revealed that *RtR* had no significant impacts on key outcomes.

More specifically, *RtR* did not have an overall impact on recent sexual activity or on sexual intercourse without birth control at either the short-term follow-up (12 months after baseline) or the longer-term follow-up (24 months after baseline). Treatment and control group youth were comparable in their engagement in sexual activity and in sexual intercourse without birth control at both follow-up time points. Moreover, as expected, sexual activity increased over time (as students aged) among both treatment and control group youth. At the short-term follow-up, slightly more than a quarter of youth in both groups were recently sexual active; at the longer-term follow-up, more than one-third of youth in both groups were recently sexually active. Engagement in sexual intercourse without birth control also increased from the short-term to the longer-term follow-up. At the short-term follow-up, 9 percent of treatment and control group youth had recently engaged in sexual intercourse without birth control; at the longer-term follow-up, 12 percent had.

RtR also had no overall impact on pregnancy in the period between baseline and the longer-term follow-up (see Exhibit 5.1). Youth in both treatment and control groups were equally likely to have gotten pregnant or gotten someone pregnant since baseline. In both groups less than 6 percent of youth reported getting pregnant (or getting someone pregnant) between baseline and the longer-term follow-up, 24 months later.

Exhibit 5.1: Short-Term and Longer-Term Impacts on Key Behavioral Outcomes

Outcome	Short-Term Impacts				Longer-Term Impacts			
	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	*p*-Value	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	*p*-Value
Sexual Behavior								
Sexual activity (percentage responding affirmatively)								
Recently sexually active (in the last 90 days)[c]	28.02	28.14	-0.11	.946[d]	35.95	34.35	1.59	.378[d]
Sexual risk behavior (percentage responding affirmatively)								
Sexual intercourse without birth control (in the last 90 days)	8.73	8.99	-0.25	.815[d]	12.09	11.64	0.45	.719[d]
Consequences of sexual risk behavior (percentage responding affirmatively)								
Pregnant or gotten someone pregnant since baseline	n/a	n/a	n/a	n/a	5.53	5.91	-0.38	.683[e]

Source: Follow-up surveys administered 12 months after baseline and 24 months after baseline.

Notes: Short-term results in this table are based on 2,665–2,667 respondents who provided valid survey responses to relevant items. Longer-term results are based on 2,720–2,780 respondents who provided valid 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 randomization 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/or anal sex. Youth were not asked about anal sex in LifeWorks or San Diego Youth Services.

[d] After application of a Benjamini-Hochberg (1995) correction for two tests within this outcome domain, the criterion for statistical significance is $p<.05$ if both tests have *p*-values less than .05, and .025 if only one of the two tests has a *p*-value less than .05.

[e] Criterion for statistical significance is $p<.05$.

5.2 Exploratory Analyses of Effects on Additional Behavioral Outcomes

In addition to the confirmatory analyses described above, we conducted a series of exploratory analyses. Though only suggestive of evidence of program effectiveness, they reflect theory (the program logic model), are supported by the experimental study design, and were specified in advance of the analysis. These exploratory analyses tested program effects on other behavioral outcomes related to sexual activity, sexual risk behavior, and consequences for the full sample (see Exhibit 5.2).

Exploratory analyses revealed that *RtR* had no overall effects on other behavioral outcomes at the short-term or longer-term follow-up.

Regarding sexual activity, engagement in recent sexual intercourse and oral sex increased between the short-term (12 months after baseline) and longer-term (24 month after baseline) follow-up, but rates for treatment and control groups were comparable. Specifically, slightly less than one-quarter of students had engaged in recent sexual intercourse at the short term, compared with about 30 percent in the longer term.

Almost 20 percent had recently engaged in oral sex at the short term, compared with one-quarter in the longer term. The percentages of treatment and control youth who had initiated sexual activity were also comparable at both time points. After 24 months, more than one-third of those who were not sexually active at baseline had initiated sexual activity.

Treatment and control group youth were equally likely to have engaged in sexual risk behavior at both the short- and longer-term follow-up. At the time of the short-term follow-up, about 15 percent of youth had recently engaged in sexual intercourse without a condom and in oral sex without a condom. At the time of the longer-term follow-up, about 20 percent of youth had.

At the longer-term follow-up, *RtR* had no effect on STI diagnoses. Less than two percent of youth in both the treatment and control groups reported being diagnosed with an STI in the past 12 months.

Exhibit 5.2: Additional Short-Term and Longer-Term Effects on Sexual Activity, Sexual Risk Behavior and Consequences

Outcome	Short-Term Impacts				Longer-Term Impacts			
	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	*p*-Value	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	*p*-Value
Sexual Behavior								
Sexual activity (percentage responding affirmatively)								
Sexual intercourse in the last 90 days	23.66	24.37	-0.72	.671	31.32	29.64	1.68	.361
Oral sex in the last 90 days	19.24	19.50	-0.26	.871	25.81	25.10	0.70	.677
Initiation of sexual activity[c]	24.98	21.96	3.02	.156	37.96	34.19	3.77	.118
Sexual risk behavior (percentage responding affirmatively)								
Sexual intercourse without a condom (in the last 90 days)	13.57	15.38	-1.81	.178	20.12	19.32	0.80	.604
Oral sex without a condom (in the last 90 days)	16.20	17.33	-1.13	.444	22.39	22.21	0.17	.912
Consequences of sexual risk behavior (percentage responding affirmatively)								
Diagnosed with STI in the last 12 months	n/a	n/a	n/a	n/a	1.66	1.81	-0.15	.777

Source: Follow-up surveys administered 12 months after baseline and 24 months after baseline.

Notes: Short-term results in this table are based on 2,661–2,667 respondents who provided valid survey responses to relevant items. Longer-term results are based on 2,777–2,780 respondents who provided valid 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 randomization 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/or anal sex. Youth were not asked about anal sex in LifeWorks or San Diego Youth Services. The sample size for the initiation of sexual activity outcome at the short-term is 1,836, as this outcome only includes youth who were not sexually active at baseline. The sample size at the longer-term is 1,932.

6. Exploratory Analyses of Program Effects on Non-Behavioral Intermediate Outcomes

RtR's theory of change (see logic model in Exhibit 2.2) specifies a set of intermediate outcomes that the model predicts will influence behavior. If the theory underlying the logic model is correct, we would expect positive effects on these non-behavioral intermediate outcomes in the short term, and that those effects would be sustained over time to change student behavior in ways that ultimately protect them from the potential consequences of sexual risk behavior (e.g., from STIs and early pregnancy).

Accordingly, the study is designed to determine whether *RtR* affects those non-behavioral outcomes Specifically, when delivered with fidelity, the program is intended to affect students' knowledge and understanding of reproductive health and avoidance of sexual risk, attitudes toward using protection, motivation to delay childbearing, intentions to become sexually active and use protection, and skills needed to avoid sexual risk.

Although the primary goals of the TPP Program (and the *RtR* program model) are to effect positive change in sexual activity, sexual risk behavior, and the consequences of that behavior, these non-behavioral outcomes remain of interest, even at the longer-term follow-up, as precursors to behavioral change. This is true particularly for the sizeable proportion of the sample who had not yet become sexually active by the end of the study (and, therefore, had not had an opportunity to demonstrate safe sexual behavior). As we noted earlier, the analyses in this chapter and the next are exploratory.

Key Non-Behavioral Findings

At both the short-term (12 months after baseline) and longer-term (24 months after baseline) follow-ups:

- *RtR* increased students' knowledge of pregnancy risk and knowledge of STI risk.
- *RtR* slightly improved students' attitudes toward using birth control or condoms.

At 24 months:

- *RtR* increased students' motivation to delay childbearing.
- *RtR* slightly increased students' confidence in their condom negotiation skills.

The program had no effect on:

- Attitudes toward risky sexual behavior.
- Motivation to delay childbearing (at 12 months).
- Intentions to engage in sexual activity or to use condoms or other birth control.
- Perceived refusal skills.

In the sections below, we briefly report on the findings related to these non-behavioral, intermediate outcomes at the short-term and longer-term follow-up.[16]

6.1 Knowledge of Pregnancy and STI Risk

At both the short-term (12 months after baseline) and longer-term (24 months after baseline) follow-up, *RtR* had statistically significant effects on knowledge of pregnancy risk and knowledge of STI risk.

Compared with control group students, treatment group students knew significantly more about pregnancy risk, including topics such as the effectiveness of condoms and birth control in preventing pregnancy. They scored about 4 percentage points higher, on average, than did control group students at both time points (Exhibit 6.1).

[16] More detail on short-term effects on intermediate outcomes can be found in the short-term report at https://aspe.hhs.gov/pdf-report/teen-pregnancy-prevention-replication-study-short-term-impacts-reducing-risk.

Students in the treatment group also scored significantly higher than did students in the control group on the measure reflecting knowledge of STI transmission and prevention, which included topics such as the effectiveness of birth control and condoms in preventing HIV or other STIs. After 12 months, the treatment group scored about 4 percentage points higher, on average. By 24 months, the gap had narrowed, but there was still a statistically significant 2 percentage point difference between the treatment and control groups.

Exhibit 6.1: Short-Term and Longer-Term Effects of *Reducing the Risk* on Knowledge

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]		*p*-Value
Short-Term Follow-Up					
Knowledge of pregnancy risk[c]	65.55	61.55	4.01	***	.000
Knowledge of STI risk[c]	60.47	56.21	4.26	***	.000
Longer-Term Follow-Up					
Knowledge of pregnancy risk[c]	68.79	64.41	4.38	***	.000
Knowledge of STI risk[c]	61.68	59.52	2.16	**	.010

Source: Follow-up surveys administered 12 months and 24 months after baseline.

Note: Results in this table are based on 2,689 respondents (short-term survey) and 2,799 respondents (longer-term survey) who provided valid survey responses to relevant items.

[a] The treatment group mean is regression adjusted, calculated as the sum of the 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 randomization blocks and other covariates. The treatment effect is expressed in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[c] Scores represent the average percentage of items answered correctly.

* $p<.05$, ** $p<.01$, *** $p<.001$ (two-tailed tests).

6.2 Attitudes

***RtR* improved students' attitudes toward using birth control or condoms.** Students in the treatment group had more positive (and protective) attitudes than did students in the control group at both the short-term and longer-term follow-ups. At both time points, however, the effect size was rather small.

The program had no statistically significant effects on student attitudes toward risky sexual behavior. Almost all students in both the treatment and control groups at both time points rejected the view that such behaviors were acceptable (Exhibit 6.2).

Exhibit 6.2: Short-Term and Longer-Term Effects of *Reducing the Risk* on Attitudes

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]		SES[c]	*p*-Value
Short-Term Follow-Up						
Attitudes toward protection[d]	3.18	3.13	0.05	***	0.13	.000
Attitudes toward risky sexual behavior[e]	5.32	4.53	0.80		n/a	.161
Longer-Term Follow-Up						
Attitudes toward protection[d]	3.16	3.13	0.03	*	0.08	.027
Attitudes toward risky sexual behavior[e]	6.00	5.52	0.49		n/a	.448

Source: Follow-up surveys administered 12 months and 24 months after baseline.

Note: Results in this table are based on 2,675-2,688 respondents (short-term survey) and 2,790–2,799 respondents (longer-term survey) who provided valid survey responses to relevant items.

[a] The treatment group mean is regression adjusted, calculated as the sum of the 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 randomization blocks and other covariates. For outcomes reported as percentages, the treatment effect is expressed 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.

[c] The SES is the standardized effect size of the difference. For outcomes that are not reported as percentages, the SES is the treatment effect divided by the pooled standard deviation of the treatment and control groups. n/a is not applicable.

[d] Scale score averages responses ranging from 1 to 4. Higher scores indicate more positive attitudes.

[e] Score represents the average percentage of items agreed with (ranging from 0 to 100). Higher values indicate more support for risky sexual behavior.

* p<.05, ** p<.01, *** p<.001 (two-tailed tests).

6.3 Motivation to Delay Childbearing

***RtR* had a positive effect on motivation to delay childbearing after 24 months only.** After 12 months, students in both the treatment and control groups were highly motivated to delay childbearing and indicated a belief in the importance of delaying childbearing until personal goals had been achieved. That is, *RtR* had no effect in the short term. After 24 months, however, control group youth declined slightly in their motivation to delay childbearing, and a small, but significant difference between treatment and control group youth emerged (Exhibit 6.3).

Exhibit 6.3: Short-Term and Longer-Term Effects of *Reducing the Risk* on Motivation to Delay Childbearing

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]		SES[c]	*p*-Value
Short-Term Follow-Up						
Motivation to delay childbearing[d]	3.68	3.68	-0.01		-0.01	.741
Longer-Term Follow-Up						
Motivation to delay childbearing[d]	3.66	3.61	0.05	*	0.09	.025

Source: Follow-up surveys administered 12 months and 24 months after baseline.

Note: Results in this table are based on 2,683 respondents (short-term survey) and 2,793 respondents (longer-term survey) who provided valid survey responses to relevant items.

[a] The treatment group mean is regression adjusted, calculated as the sum of the 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 randomization 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 SES is the standardized effect size of the difference. The SES is the treatment effect divided by the pooled standard deviation of the treatment and control groups.

[d] Scale score averages responses ranging from 1 to 4. Higher scores indicate higher motivation.

* p<.05, ** p<.01, *** p<.001 (two-tailed tests).

6.4 Intentions

***RtR* did not affect students' intentions to engage in sexual activity.** At both time points, similar proportions of students in the treatment and control groups expected to engage in sexual intercourse or oral sex in the next 12 months. **Neither did the program affect students' intention to use condoms or**

birth control if they were to engage in sexual intercourse. At both time points, nearly all students (about 90 percent) reported that they intended to do so (Exhibit 6.4).

Exhibit 6.4: Short-Term and Longer-Term Effects of *Reducing the Risk* on Intentions

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	*p*-Value
Short-Term Follow-Up				
Sexual intercourse [c]	52.67	50.69	1.97	.280
Oral sex [c]	42.41	43.27	-0.86	.632
Use birth control if they were to have sexual intercourse [c]	90.39	89.67	0.72	.537
Use a condom if they were to have sexual intercourse [c]	91.21	92.11	-0.90	.403
Longer-Term Follow-Up				
Sexual intercourse [c]	60.47	58.68	1.79	.331
Oral sex [c]	50.36	51.78	-1.43	.406
Use birth control if they were to have sexual intercourse [c]	89.52	88.75	0.77	.523
Use a condom if they were to have sexual intercourse [c]	89.90	88.96	0.93	.447

Source: Follow-up surveys administered 12 months and 24 months after baseline.

Note: Results in this table are based on 2,654–2,667 respondents (short term survey) and 2,764–2,781 respondents (longer-term survey) who provided valid survey responses to relevant items.

[a] The treatment group mean is regression-adjusted, calculated as the sum of the 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 randomization blocks and other covariates. The treatment effect is expressed in percentage points. Due to rounding, reported treatment effects may differ from differences between reported means for the treatment and control groups.

[c] Outcomes measure intention to engage in the behavior in the next 12 months. Dichotomous variables, reported as percentage of respondents who responded affirmatively.

6.5 Skills

***RtR* had a favorable effect on condom negotiation skills after 24 months only.** After 12 months, *RtR* had no significant effect on either perceived refusal skills or perceived condom negotiation skills. Treatment and control group students perceived their abilities on both measures to be equally strong. After 24 months, treatment group youth were slightly more certain of their condom negotiation skills than were control group youth. That is, in the longer term, the program had a significant positive effect (Exhibit 6.5).

Exhibit 6.5: Short-Term and Longer-Term Effects of *Reducing the Risk* on Skills

Outcome	Adjusted Treatment Mean[a]	Unadjusted Control Mean	Treatment Effect[b]	SES[c]	*p*-Value
Short-Term Follow-Up					
Perceived refusal skills[d]	3.12	3.08	0.04	0.06	.132
Perceived condom negotiation skills[d]	3.53	3.50	0.03	0.06	.177
Longer-Term Follow-Up					
Perceived refusal skills[d]	3.18	3.15	0.03	0.04	.263
Perceived condom negotiation skills[d]	3.53	3.49	0.04 *	0.08	.030

Source: Follow-up surveys administered 12 months and 24 months after baseline.

Note: Results in this table are based on 2,681–2,685 respondents (short-term survey) and 2,793–2,794 respondents (longer-term survey) who provided valid survey responses to relevant items.

[a] The treatment group mean is regression adjusted, calculated as the sum of the 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 randomization 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 SES is the standardized effect size of the difference. The SES is the treatment effect divided by the pooled standard deviation of the treatment and control groups.

[d] Scale score averages responses ranging from 1 to 4. Higher scores indicate greater certainty about skills.

* $p<.05$, ** $p<.01$, *** $p<.001$ (two-tailed tests).

7. Exploratory Analyses of Program Effects by Site and Subgroup

The results of the confirmatory analyses reported in Chapter 5 offer the best evidence to answer with confidence whether or not *RtR* had an impact. At the same time, the amount of data collected and pooled across the three sites allowed us to conduct additional exploratory analyses related to possible variation in effects by site or for certain subgroups. The results of those analyses, presented here, must be interpreted with caution and primarily be viewed as hypothesis generating, rather than as additional conclusive evidence on program impacts. The reason for this caution is simple: The large number of tests conducted in these exploratory analyses increases the risk of producing a significant finding simply by chance, and no adjustments are made to reduce that risk. We cautiously interpret findings in cases where we can identify a pattern of either positive or negative findings in the same direction.

With this caveat, in this chapter we present the results of analyses that examined site-level differences in effects on the behavioral outcomes and non-behavioral intermediate outcomes described in the previous chapters. We also examine differences in effects on these same outcomes for different subgroups based on gender, age, race/ethnicity, and baseline sexual experience. Again, we emphasize that readers should note that large numbers of tests were conducted in these analyses and we would expect for 5 percent of the tests conducted to see statistically significant test results even when there were no real effect of the intervention.

Key Site and Subgroup Findings

Site-Level Differences:

- *RtR* had significant effects, both favorable and unfavorable, on behaviors for some sites.
- There were some site-level differences in effects on non-behavioral intermediate outcomes at the short-term, but not the longer-term follow-up.

Subgroup Differences

- At the longer-term follow-up only, there were significant differences in the effects of *RtR* on recent sexual activity and oral sex based on race/ethnicity.
- At the short-term follow-up, there were significant differences in the effects of *RtR* on attitudes toward risky sexual behavior based on age and race/ethnicity.
- At the longer-term follow-up, there were subgroup differences (based on age and gender) in the effects of *RtR* on knowledge of pregnancy risk, attitudes toward protection, and perceived condom negotiation skills.

Tables documenting the site-level findings can be found in Appendix B, and the corresponding tables documenting subgroup findings are in Appendix C.

7.1 Site-Level Differences

In this section we discuss findings related to site-level differences in effects on both behavioral and non-behavioral intermediate outcomes. We tested for site-level differences in these effects at both the short-term and longer-term follow-up periods.

7.1.1 Behavioral Outcomes

Though there were no significant program impacts for the pooled sample, **there were significant effects, both favorable and unfavorable, on behaviors for some sites.** Exploratory site-level analyses revealed significant differences in effects on behavior by site at both the short- and longer-term follow-ups.

At the short-term follow-up, *RtR* had a significant and favorable effect on students in the BFL sample, but not in the other two sites (Appendix Exhibit B.1). That is, in BFL, significantly 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).

At the longer-term follow-up, there were significant site-level differences in program effects on sexual activity outcomes (see Appendix Exhibit B.2). Though in the intended direction, the effect on recent sexual intercourse in BFL was no longer significant at the longer-term follow-up. However, there was a favorable effect on recent engagement in oral sex in BFL. That is, program participants were less likely (5.9 percentage points) to have engaged in oral sex in the last 90 days than were their control group counterparts.

By contrast, at the longer-term follow-up, there were unintended effects on sexual activity in LifeWorks. Specifically, program participants in this site were more likely (8.6 percentage points) to have engaged in sexual activity in the last 90 days than were students in the control group. Significantly more program participants (42.2 percent) reported engaging in sexual intercourse in the 90 days before the survey compared with their control group counterparts (33.9 percent). *RtR* participants in LifeWorks were also more likely (6.5 percentage points) to have engaged in oral sex in the last 90 days than were students in the control group. There were no effects of *RtR* on behavior in SDYS.

At the longer-term follow-up, the effect of *RtR* on pregnancy also varied by site (see Appendix Exhibit B.2). Specifically, in BFL, fewer program participants (8.3 percent) reported that they had become pregnant or had gotten someone pregnant since baseline compared with students in the control group (12.2 percent). There were no program effects on pregnancy in the other two sites.

7.1.2 Non-Behavioral Intermediate Outcomes

A few significant differences between sites in the effect of *RtR* on non-behavioral intermediate outcomes emerged at the short-term follow-up (see Appendix Exhibit B.3). Significant site-level differences in effects on attitudes toward protection and perceived condom negotiation skills revealed a favorable effect in SDYS, but not in the other two sites. In SDYS, students in the treatment group had significantly more positive attitudes toward protection (SES = 0.30) compared with their control group counterparts, suggesting that this effect in SDYS likely drove the significant effect found for the full sample. Though there were no effects on skills found for the full sample at the short-term follow-up, *RtR* participants in SDYS reported stronger perceived condom negotiation skills (SES = 0.19) compared with SDYS students in the control group.

Although there were no effects found for the full sample on motivation or intentions at the short-term follow-up, significant site-level differences in short-term effects on motivation to delay childbearing and intentions to have oral sex revealed effects in BFL, but not in the other two sites. In BFL, there was an unintended effect on motivation to delay childbearing. That is, students in the treatment group were less motivated to delay childbearing compared with their control group counterparts (SES = −0.18). However, there was a favorable effect in BFL on intentions to have oral sex. Students in the treatment group were less likely than their counterparts in the control group (7.4 percentage points) to intend to have oral sex in the 12 months after the short-term survey.

At the longer-term follow-up, none of these site-level differences in effects on non-behavioral intermediate outcomes persisted (see Appendix Exhibit B.4).

7.2 Subgroup Differences

We also conducted exploratory analyses to look at differences in program effects on behavioral and non-behavioral intermediate outcomes by student subgroup. Specifically, we looked at whether program effects differed by gender, age, race/ethnicity, or baseline sexual experience at the short-term and longer-

term follow-ups. We present impact estimates for individual subgroups only when there is a statistically significant difference in program effect between subgroups.

7.2.1 Behavioral Outcomes

At the short-term follow-up, exploratory analyses revealed no significant differences in the effects of *RtR* on sexual activity or sexual risk by subgroup (i.e., based on age, gender, race/ethnicity, or sexual experience at baseline). However, **at the longer-term follow-up, there were significant differences between racial/ethnic groups** in the behavioral effects of *RtR*. There were unintended effects of *RtR* on recent sexual activity and oral sex in the last 90 days for Hispanic youth. Hispanic students in the treatment group were more likely to report engaging in sexual activity in the last 90 days (6.8 percentage points more likely) and more specifically, engaging in oral sex in the last 90 days (5.8 percentage points more likely) than were Hispanic students in the control group. There were no significant effects on either of these two behavioral outcomes for other racial/ethnic groups (see Appendix Exhibit C.1).

It is important to note that these unfavorable program effects among Hispanic youth parallel the unfavorable effects found for the LifeWorks site. Given that the LifeWorks site largely consisted of Hispanic youth, we conducted additional analyses to determine whether these unfavorable effects were driven by the site or by the Hispanic subgroup (see Appendix F). We conclude that **there were unique unfavorable effects both for Hispanic youth and for youth in the LifeWorks site**; neither effect drove the other.

7.2.2 Non-Behavioral Intermediate Outcomes

Though there were no overall program effects on attitudes toward risky sexual behavior **at the short-term follow-up, there were significant differences in the effects of *RtR* on attitudes toward risky sexual behavior based on age and race/ethnicity, suggesting adverse program effects for older students and for White students** (see Appendix Exhibit C.2). Older, but not younger, students who received the program had attitudes that were more supportive of risky sexual behavior compared with older students in the control group (2.3 percentage point difference on support scale). Similarly, White program participants had attitudes that were more supportive of risky sexual behavior compared with White students in the control group (5.7 percentage point difference on support scale). There were no effects on attitudes toward risky sexual behavior for Hispanic, Black, or Other race/ethnicity subgroups.

These significant differences in effects on attitudes toward risky sexual behavior based on age and race/ethnicity did not persist to the 24-month follow-up. However, **at 24 months, significant subgroup differences did emerge in the effects of *RtR* on knowledge of pregnancy risk, on attitudes toward protection, and on perceived condom negotiation skills** (see Appendix Exhibit C.3). For older, but not younger youth, *RtR* had a positive effect on knowledge of pregnancy risk: Older youth in the treatment group scored 6.4 percentage points higher on the knowledge scale, on average, compared with older control group youth, suggesting it is older youth that drive the program effect on knowledge of pregnancy risk found for the full sample. For male, but not female, students, *RtR* had a positive effect on attitudes toward protection and on perceived condom negotiation skills. Male students in the treatment group reported more positive attitudes toward protection and stronger perceived condom negotiation skills compared with their control group counterparts (treatment-control difference was 0.1 on a four-point scale for both outcomes). These subgroup findings qualify the effects found for the full sample, suggesting that

it is the effects among male students that drives the overall effects found for attitudes toward protection and perceived condom negotiation skills.

7.3 Differences in Program Effects on Youth Based on Sexual Activity Throughout the Study

Our earlier discussion about possible different prevention and intervention program effects (see Section 2.2) raised a question about whether *RtR* was differentially effective in changing the behavior of those youth who were not sexually experienced at baseline but became sexually active in the course of the study compared with youth who were already sexually experienced at the start of the study.

It is important to recognize that, to varying extents, in all three replications, ***RtR* served as both a prevention and an intervention program**:

- With youth who were sexually inexperienced when the study began, the program could have functioned as a **prevention** program: For *youth who became sexually active during the study*, the program had the potential to affect non-behavioral intermediate outcomes, which could have then translated into favorable impacts on sexual behavior (i.e., preventing risky sexual behavior). For *youth who never became sexually active during the study*, and would not have become sexually active in the absence of the program, the best the program could have demonstrated was positive and sustained effects on the non-behavioral intermediate outcomes that the logic model suggests are the foundation for later positive behavioral outcomes.
- For *youth who were sexually experienced when they entered the study*, the program could have functioned as an **intervention** program and could have had effects on non-behavioral outcomes, as well as favorable impacts on behavioral outcomes by intervening to reduce recent sexual activity and sexual risk behavior and consequences.

To begin to explore potential differences based on youth sexual activity throughout the study, and to begin to disentangle possible prevention and intervention effects, we plotted the means of the key behavioral outcomes (i.e., sexual activity in the last 90 days; sexual intercourse without birth control in the last 90 days) at each time point (baseline and 12 months and 24 months after baseline) for youth in all three subgroups.

These graphs (see Appendix D, Exhibit D.1) suggest that, though there are different trajectories of sexual behavior for the three groups, **there are no noticeable treatment-control differences between the group of youth who were sexually active when the study started and the group who initiated sexual activity during the study period.** In this preliminary exploration, we therefore did not see different prevention and intervention effects. We also looked descriptively at graphs for individual sites and did not notice consistent differences.

Because these subgroups were defined based on events that occurred after random assignment (i.e., became or did not become sexually active after the study began), fully answering this question means moving beyond the experimental framework of the study, and thus is beyond the scope of this report.

8. Discussion

The TPP Replication Study represents a strong evaluation of the *RtR* program model. The study applied a stringent test to three replications of the same program, implemented by different grantees, in different geographic locations, with quite different youth populations,[17] but within the same timeframe and using identical and comprehensive outcome measures.

In all three replications, the program was implemented with fidelity to the core elements of the model and, in all three, attendance levels were high, meaning that program participants were adequately exposed to the content.

We found no evidence that *RtR* had an overall impact on behavior across multiple implementations of the program, although there was suggestive evidence of favorable impacts in one site.

Though *Reducing the Risk* significantly improved and sustained students' knowledge and attitudes toward protection, our confirmatory analyses revealed that these positive effects on non-behavioral intermediate outcomes did not translate into significant favorable impacts on the key behavioral outcomes selected to represent the primary goals of this and all other TPP programs. Nor did the program produce favorable effects on other measures of sexual activity, sexual risk behavior, and adverse consequences.

What this suggests for policymakers and local agency staff is that the original evidence on the effectiveness of *RtR* provides limited guidance on the likely effectiveness of the program in different locations or with certain populations.

The program had significant, but different effects in two of the three implementation sites and on one racial/ethnic group.

Though they do not modify the overall conclusion, the analyses conducted to explore differences in program effectiveness at the site level and for different subgroups produced some suggestive findings. These exploratory analyses revealed a pattern of favorable effects over time in one site. However, in another site and for a single subgroup, a pattern of unfavorable program effects emerged.

In BFL, short-term behavioral and non-behavioral effects appear to have translated into favorable effects on (certain) sexual behaviors and on consequences at the longer-term follow-up, as the program model predicts. Most notably, students in BFL who were assigned to *RtR* were less likely than control group students to have had recent sexual intercourse at the short-term follow-up, and at the longer-term follow-up, they were less likely to have become pregnant or gotten someone pregnant. In addition, favorable effects on intentions to engage in oral sex at the short-term follow-up translated into favorable effects on recent oral sex at the longer-term follow-up. At both the short- and longer-term follow-ups, effects on other sexual activity and sexual risk outcomes for BFL, although not always significant or significantly different from the other replication sites, were consistently in the intended direction. The number and consistency of favorable effects lead us to believe that those effects were not by chance, but rather that the replication of *RtR* in BFL was successful in achieving positive behavioral outcomes.

[17] There was significant variation by age and ethnicity across the three sites. The sample also differed from that of the original study (Kirby et al., 1991), where the sample was primarily White.

By contrast, in LifeWorks we see unintended effects on sexual activity in the longer term that were preceded by a pattern of unintended trends at the short-term follow-up. Although these unfavorable short-term trends were not statistically significant, they assume some importance in light of the longer-term unintended effects. After 24 months, students assigned to *RtR* in LifeWorks were more likely to be recent sexual activity, to have had sexual intercourse in the last 90 days, and to have had oral sex in the last 90 days compared with students in the control group. Impacts on sexual risk behaviors and on pregnancies, though not statistically significant, were also consistently in the unintended direction in LifeWorks.

Exploratory analyses also revealed an unintended effect on sexual activity among Hispanic students. Those assigned to *RtR* were more likely to be recently sexually active and, more specifically, to have engaged in oral sex in the last 90 days, compared with their control group counterparts. It is tempting to conclude that the two sets of unintended findings are really the same effect, as the majority of study participants in Lifeworks were Hispanic. However, additional analyses (see Appendix F) suggest that there are unique effects of ethnicity and of site.

The search for plausible explanations for these findings did not reveal obvious answers. In each of the three sites, the program was implemented with fidelity and program attendance was relatively high. Despite the length of the curriculum, grantees and schools were able to accommodate the program within the class/school schedule. Focus group reports from the implementation study suggest that the program was well received and that participants were engaged. Data obtained through interviews with teachers suggest that the combination of information and strategies was a good fit for the students. There were, as noted earlier, differences in the youth population served in the three sites (e.g., differences based on race/ethnicity, sexual activity, and risk behavior at baseline), but those differences cannot readily be used to explain the differences in outcomes.

We also have not found a clear explanation for the program's negative effects on Hispanic students. To our knowledge, this is the first test of *RtR* that includes a large sample of Hispanic youth. Though the program is intended for youth of all ethnicities, it is possible that this idiosyncratic pattern of findings reflects a lack of fit between the program and a specific ethnic group.

What the conflicting findings for different sites and subgroups suggest is that context matters. It seems likely that interactions among the population served (in terms of both demographic characteristics and cultural beliefs), the attitudes and beliefs of the staff delivering the curriculum, and characteristics of the school and classroom settings influenced the program's impact in complex ways. Additional research is needed to achieve a better understanding of these interactions and their effects to help elucidate when *RtR* might work and when it might not.

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 effectiveness of *Reducing the Risk*. **Based on this study, we cannot conclude that *Reducing the Risk* was effective. Exploratory analyses suggest that the program may be beneficial in some cases, but may lead to unintended effects in other cases.**

References

Barbee, A. P., Cunningham, M. R., van Zyl, M. A., Antle, B. F., & Langley, C. N. (2016). Impact of two adolescent pregnancy prevention interventions on risky sexual behavior: A three-arm cluster randomized control trial. *American Journal of Public Health,* 106, S85-S90.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300.

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. J*ournal of Adolescent Health, 54*(5), 499-507.

Kappeler, E., & 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.

Kelsey, M., Blocklin, M., Layzer, J., Price, C., Juras, R., & Freiman, L. (2016). Reducing the Risk: 12-month impact findings of a cluster randomized controlled trial. *American Journal of Public Health, 106*(S1), S45-S52.

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.

Reyna, V.F., & Mills, B.A. (2014). Theoretically motivated interventions for reducing sexual risk taking in adolescence: A randomized controlled experiment applying fuzzy-trace theory. *Journal of Experimental Psychology. General, 143*(4), 1627-1648.

Schochet P.Z. (2008a). *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations* (NCEE 2008-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schochet, P.Z. (2008b). 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.* Washington, DC: U.S. Department of Health and Human Services, Office of Adolescent Health, Office of Public Health and Science.

Zimmerman, R.S., Cupp, P.K., Donohew, L., Sionean, C.K., Feist-Price, S., & Helme, D. (2008). Effects of a school-based, theory-driven HIV and pregnancy prevention curriculum. *Perspectives on Sexual and Reproductive Health, 40*(1), 42-51.

Appendix A: Measures

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

- Youth sexual activity, sexual risk behavior, and consequences of sexual risk behavior.
- Non-behavioral intermediate outcomes.

Measures of *youth sexual activity, sexual risk behavior,* and *consequences of sexual risk behavior* include recent sexual activity, sexual intercourse, and oral sex, sexual initiation, recent sexual intercourse without birth control, sexual intercourse without a condom, and oral sex without a condom, and pregnancy and STI diagnoses. Measures of *non-behavioral 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. In the sections that follow, we describe each category by defining constituent measures and their construction.

A.1 Youth Sexual Activity, Sexual Risk Behavior, and Consequences of Sexual Risk Behavior

To understand program effects on youths' sexual activity, sexual risk behavior, and consequences of sexual risk behavior, we examined the nine items presented in Exhibit A.1.

Exhibit A.1: Youth Sexual Activity, Sexual Risk Behavior, and Sexual Consequences Measures

Measure	Item	Coding
Sexual Behavior Outcomes		
Sexual Activity		
Recent sexual activity (in the 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/or 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.

Measure	Item	Coding
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.
Sexual Risk Behavior		
Sexual intercourse without birth control (in the 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.
Sexual intercourse without a condom (in the 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.
Oral sex without a condom (in the 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 Consequences (Longer-term follow-up only)		
Pregnant or gotten someone pregnant since baseline	To the best of your knowledge, have you ever been pregnant, or gotten someone pregnant, even if no baby was born?	This outcome measure was coded as 1=yes, 0=no indicating whether or not respondents reported that they had been pregnant or gotten someone pregnant between baseline and the longer-term follow up. When youth reported a greater number of pregnancies at the longer-term survey than at baseline, the youth was assigned a score of 1. Youth who reported the same number at baseline and the longer-term follow-up were assigned a score of 0.
Diagnosed with STI in the last 12 months	In the past 12 months, have you been told by a doctor or nurse that you had a sexually transmitted disease (STD)/ sexually transmitted infection (STI) or HIV?	Youth responded to this question with a yes(1)/no(0) answer.

A.2 Non-Behavioral Intermediate Outcomes

Non-behavioral intermediate outcomes are those expected to lead to changes in behavior. 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 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 A.2.

- ***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 A.2 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 (multiplied by 100) of 12 binary variables pertaining to knowledge of STI prevention, transmission, and treatment (See Exhibit A.2 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 A.2: 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.

Component Items	Coding
Knowledge of STI Risk (12 items)	
You can't get infected with HIV if you have sex only once or twice without 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.
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 and longer-term surveys 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 sexual behavior*. These measures are described below and detailed information about their component items is presented in Exhibit A.3.

- ***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 A.3 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 demonstrated acceptable internal consistency reliability ($\alpha = 0.75$). [18]

- ***Attitudes toward risky sexual behavior*** is a composite measure that is the mean of seven binary items (multiplied by 100) querying the acceptability and normativity of risky sexual behaviors. (See Exhibit A.3 for coding and other details.) Scores on this scale range from 0 to 100 and represent the percentage of items agreed with: Higher values reflect more support for risky sexual behavior. The measure demonstrated good internal consistency reliability ($\alpha = 0.82$).

Exhibit A.3: 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.

[18] 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."

Component Items	Coding
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.
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 and longer-term surveys 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 A.4 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 A.4: 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 A.5 to examine impacts on youth's intended or anticipated sexual activity and sexual risk behavior in the coming year.

Exhibit A.5: 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 and longer-term follow-up surveys 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 A.6.

- ***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 A.6 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 demonstrated 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 A.6 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 demonstrated good internal consistency reliability ($\alpha = 0.83$).

Exhibit A.6: 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 without 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.

Appendix B: Site-Level Effects

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 analysis of the pooled data is the primary focus of this study, there was clearly considerable interest on the part of study stakeholders in examining effects in 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, 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 effects are discussed in the main text, as we only interpret site-specific effects when a significant difference between sites is found.

Exhibit B.1: Short-Term Effects on Sexual Activity and Sexual Risk Behavior 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 activity (percentage responding affirmatively)													
Recently sexually active (in last 90 days)[d]	38.35	41.37	-3.02	.285	35.29	31.13	4.16	.160	10.99	12.30	-1.31	.672	.193
Sexual intercourse in the last 90 days	32.71	39.34	-6.63 *	.015	30.87	25.74	5.13	.073	8.65	8.62	0.03	.992	.011 *
Oral sex in the last 90 days	23.06	23.29	-0.23	.932	26.82	25.55	1.27	.647	7.44	9.42	-1.98	.497	.722
Sexual risk behavior (percentage responding affirmatively)													
Sexual intercourse without birth control (in last 90 days)	11.20	13.11	-1.91	.300	12.33	11.03	1.30	.493	2.83	2.87	-0.04	.984	.475
Sexual intercourse without a condom (in last 90 days)	17.47	22.13	-4.66 *	.037	19.65	19.85	-0.20	.933	3.93	4.18	-0.25	.918	.282
Oral sex without a condom (in last 90 days)	17.89	19.45	-1.56	.529	23.97	23.83	0.14	.956	6.35	8.38	-2.03	.455	.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 randomization 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.

* $p<.05$, ** $p<.01$, *** $p<.001$ (two-tailed tests).

Exhibit B.2: Longer-Term Effects on Sexual Activity, Sexual Risk Behavior, and Sexual Consequences by Site

Outcome	Better Family Life (n= 854)				LifeWorks (n= 894)				San Diego Youth Services (n= 1,033)				*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													
Sexual activity (percentage responding affirmatively) [d]													
Recently sexually active (in last 90 days)	42.85	47.71	-4.86	.121	46.87	38.30	8.57 **	.005	21.49	20.76	0.73	.805	.008 **
Sexual intercourse in the last 90 days	38.02	42.81	-4.79	.126	42.16	33.94	8.22 **	.007	17.25	15.85	1.40	.648	.012 *
Oral sex in the last 90 days	26.71	32.62	-5.91	.043	36.65	30.11	6.54 *	.021	15.88	14.73	1.15	.680	.009 **
Sexual risk behavior (percentage responding affirmatively)													
Sexual intercourse without birth control (in last 90 days)	12.07	14.68	-2.61	.239	18.75	14.91	3.84	.071	6.18	6.25	-0.07	.973	.105
Sexual intercourse without a condom (in last 90 days)	23.76	24.46	-0.70	.797	28.92	25.46	3.46	.193	9.16	9.60	-0.44	.868	.467
Oral sex without a condom (in last 90 days)	21.87	25.30	-3.43	.218	32.86	28.51	4.35	.107	13.25	13.84	-0.59	.825	.126
Sexual consequences (percentage responding affirmatively)													
Pregnant or gotten someone pregnant since baseline	8.34	12.15	-3.81*	.024	7.35	5.15	2.20	.174	2.17	2.06	0.11	.942	.034 *
Diagnosed with STI in the last 12 months	3.38	3.94	-0.56	.557	1.92	1.38	0.54	.554	0.24	0.67	-0.43	.620	.651

Source: Follow-up survey administered 24 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 randomization blocks and other covariates. The treatment effect is expressed as a difference in percentage points

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

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests).

Exhibit B.3: Short-Term Effects on Non-Behavioral Intermediate Outcomes 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	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]	
Knowledge																
Knowledge of pregnancy risk [e]	59.34	55.96	3.38	.062		71.91	69.53	2.38	.204		64.70	58.38	6.32 ***	.001		.303
Knowledge of STI risk [e]	59.15	54.52	4.63 ***	.001		63.75	60.46	3.29 *	.018		58.14	53.29	4.85 ***	.001		.687
Attitudes																
Attitudes toward protection [f]	3.20	3.19	0.01	.605	0.03	3.18	3.14	0.04	.144	0.09	3.17	3.05	0.12 ***	.000	0.30	.006 **
Attitudes toward risky sexual behavior [g]	3.88	2.80	1.08	.259		7.13	5.97	1.16	.246		4.77	4.65	0.12	.903		.713
Motivation																
Motivation to delay childbearing [f]	3.64	3.75	-0.10 *	.014	-0.18	3.71	3.66	0.04	.354	0.07	3.70	3.65	0.05	.256	0.09	.018 *
Intentions to engage in the following behaviors in the next 12 months																
Intention to have sexual intercourse in the next 12 months [h]	58.96	62.64	-3.68	.220		62.29	57.25	5.04	.108		37.72	32.38	5.34	.103		.062
Intention to have oral sex in the next 12 months [h]	38.89	46.28	-7.39 *	.011		53.95	52.58	1.37	.650		34.85	30.45	4.40	.158		.015 *
Intention to use birth control if they were to have sexual intercourse in the next 12 months [h]	89.64	90.63	-0.99	.618		90.20	88.73	1.47	.473		91.58	89.76	1.82	.378		.560
Intention to use a condom if they were to have sexual intercourse in the next 12 months [h]	91.81	93.70	-1.89	.300		89.64	92.89	-3.25	.084		92.30	89.76	2.54	.180		.076
Skills																
Perceived refusal skills [f]	3.10	3.09	0.02	.709	0.02	3.14	3.05	0.09	.062	0.12	3.12	3.10	0.02	.700	0.03	.472
Perceived condom negotiation skills [f]	3.61	3.66	-0.05	.182	-0.09	3.52	3.48	0.04	.234	0.08	3.47	3.37	0.10 **	.008	0.19	.015 *

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 randomization 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] Scores represent the average percentage of items answered correctly.

[f] Scale score averages responses ranging from 1 to 4. Higher scores indicate higher levels of the outcome.

[g] Score represents the average percentage of items agreed with.

[h] Dichotomous variables, reported as percentage of respondents who responded affirmatively.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests).

Exhibit B.4: Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Site

Outcome	Better Family Life (n= 857)					LifeWorks (n= 901)					San Diego Youth Services (n= 1,041)					*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]	
Knowledge																
Knowledge of pregnancy risk [e]	60.99	59.24	1.75	.399		74.66	71.15	3.51	.083		69.41	61.59	7.82***	.000		.098
Knowledge of STI risk [e]	58.59	58.38	0.21	.890		65.85	63.03	2.82	.052		60.19	56.94	3.25*	.022		.287
Attitudes																
Attitudes toward protection [f]	3.18	3.13	0.05	.095	0.11	3.14	3.15	-0.01	.739	-0.02	3.17	3.10	0.07**	.009	0.16	.108
Attitudes toward risky sexual behavior [g]	3.57	4.60	-1.03	.377		8.74	6.87	1.87	.096		5.38	4.86	0.52	.634		.201
Motivation																
Motivation to delay childbearing [f]	3.63	3.57	0.06	.174	0.09	3.68	3.64	0.03	.462	0.05	3.67	3.60	0.07	.084	0.11	.791
Intentions to engage in the following behaviors in the next 12 months																
Intention to have sexual intercourse in the next 12 months [h]	66.20	67.88	-1.68	.608		67.35	65.60	1.75	.585		50.15	45.05	5.10	.110		.334
Intention to have oral sex in the next 12 months [h]	50.77	54.74	-3.97	.203		59.69	60.23	-0.54	.857		41.20	41.31	-0.11	.970		.617
Intention to use birth control if they were to have sexual intercourse in the next 12 months [h]	89.58	88.07	1.51	.493		89.01	90.41	-1.40	.506		89.76	87.61	2.15	.285		.440
Intention to use a condom if they were to have sexual intercourse in the next 12 months [h]	91.96	89.06	2.90	.186		87.71	89.27	-1.56	.461		90.15	88.59	1.56	.449		.319
Skills																
Perceived refusal skills [f]	3.17	3.11	0.06	.227	0.08	3.23	3.19	0.04	.401	0.05	3.16	3.16	0.00	.959		.633
Perceived condom negotiation skills [f]	3.59	3.61	-0.02	.583	-0.04	3.53	3.46	0.08*	.032	0.14	3.50	3.43	0.07*	.041	0.13	.112

Source: Follow-up survey administered 24 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 randomization 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] Scores represent the average percentage of items answered correctly.

[f] Scale score averages responses ranging from 1 to 4. Higher scores indicate higher levels of the outcome.

[g] Score represents the average percentage of items agreed with.

[h] Dichotomous variables, reported as percentage of respondents who responded affirmatively.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests).

Appendix C: Subgroup Effects

To better understand what works for whom, we estimated effects for key subgroups of participants (based on age, gender, race/ethnicity, and sexual experience at baseline) and tested for differences in effects between subgroups. To guard against potential overinterpretation of results, we present impact estimates for individual subgroups only when there is a statistically significant difference between subgroups. For example, the impact estimate would be presented for the subgroup of boys only if there were a statistically significant difference between the effects on boys and girls.

At the **short-term follow-up,** there were no differences in effects on sexual activity or sexual risk behavior by any of the subgroups examined. The exhibits below display where differences in effects were found.

Exhibit C.1: Longer-Term Effects on Sexual Behavior by Subgroup

	Treatment Effect [a]	*p*-Value [b]
Recently sexually active (in last 90 days) (percentage responding affirmatively) [c]		
Subgroup: Respondent Race		
Hispanic (n= 1,300)	6.81**	.007
Black (n= 904)	-2.11	.484
White (n= 310)	-4.98	.317
Other (n= 264)	-3.33	.540
Oral sex in the last 90 days (percentage responding affirmatively)		
Subgroup: Respondent Race		
Hispanic (n= 1,299)	5.77*	.013
Black (n= 905)	-3.60	.195
White (n= 309)	-2.53	.581
Other (n= 264)	-5.36	.285

Source: Follow-up survey administered 24 months after baseline.

Note: Impact estimates for subgroups are shown only if a test for differences in effects among the subgroups met the study criterion for statistical significance ($p<.05$). For example, a test result indicated that the treatment effect on recent sexual activity was significantly different across racial/ethnic groups.

[a] This column shows the estimated treatment effect (treatment/control difference) 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.

[c] 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.

* $p< .05$, ** $p< .01$, *** $p< .001$ (two-tailed tests).

Exhibit C.2: Short-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup

	Treatment Effect [a]	*p*-Value [b]
Attitudes toward Risky Sexual Behavior [c]		
Subgroup: Respondent Age		
Respondent less than age 15 (n =1,545)	-0.30	.690
Respondent age 15 or older (n =1,130)	2.31**	.008
Subgroup: Respondent Race [d]		
Hispanic (n =1,170)	-0.71	.401
Black (n =960)	1.18	.214
White (n =287)	5.66***	.001
Other (n =258)	0.66	.717
Motivation to Delay Childbearing [e]		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n =1,847)	-0.05	.142
Ever sexually active at baseline (n =836)	0.07	.112

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

Note: Impact estimates for subgroups are shown only if a test for differences in effects between the subgroups met the study criterion for statistical significance (p<.05). For example, a test result indicated that the treatment effect on attitudes toward risky sexual behavior was significantly different for younger versus older respondents.

[a] This column shows the estimated treatment effect (treatment-control difference) 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.

[c] Scores represent the average percentage of items agreed with.

[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] Scale score averages responses ranging from 1 to 4. Higher scores indicate higher levels of the outcome.

* p< .05, ** p< .01, *** p< .001 (two-tailed tests).

Exhibit C.3: Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup

	Treatment Effect [a]	*p*-Value [b]
Knowledge of Pregnancy Risk [c]		
Subgroup: Respondent age		
Respondent less than age 15 (n= 1,681)	6.52***	.000
Respondent age 15 or older (n= 1,118)	1.44	.411
Attitudes toward Protection [d]		
Subgroup: Respondent gender		
Male (n= 1,391)	0.07**	.002
Female (n= 1,408)	0.00	.941
Perceived Condom Negotiation Skills [d]		
Subgroup: Respondent gender		
Male (n= 1,386)	0.10***	.000
Female (n= 1,407)	-0.01	.609

Source: Follow-up survey administered 24 months after baseline.

Note: Impact estimates for subgroups are shown only if a test for differences in impacts between the subgroups met the study criterion for statistical significance (p<0.05). For example, a test result indicated that the treatment effect on knowledge of pregnancy risk was significantly different for younger versus older respondents.

[a] This column shows the estimated treatment effect (treatment-control difference) 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.

[c] Scores represent the average percentage of items answered correctly.

[d] Scale score averages responses ranging from 1 to 4.Higher scores indicate higher levels of the outcome.

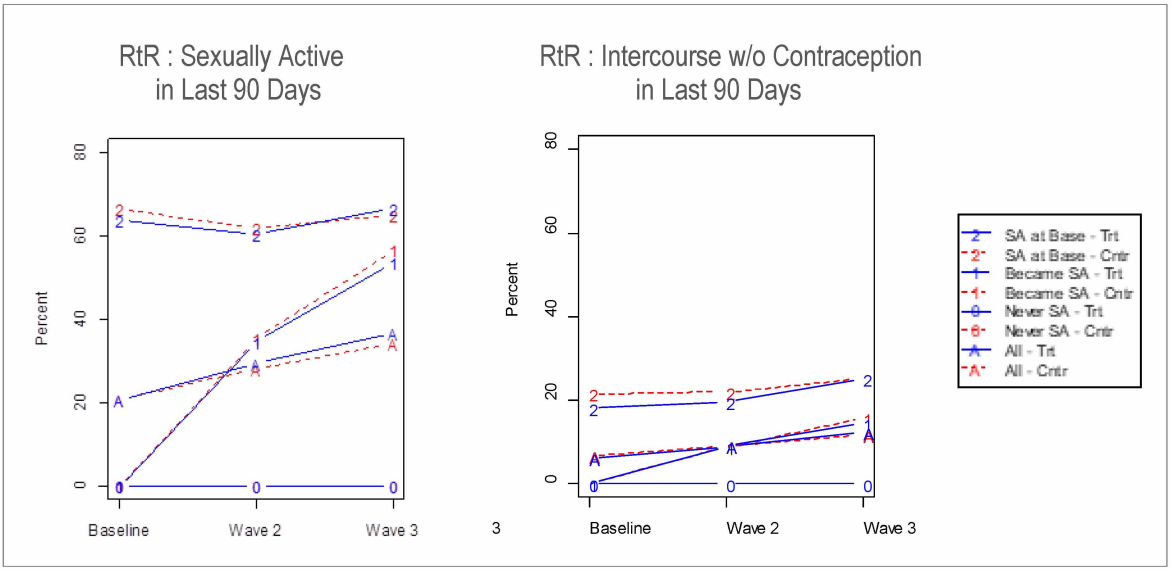
* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests).

Appendix D: Preliminary Descriptive Findings on Subgroups of Youth Based on Sexual Activity Throughout the Study

As described in the body of the report (see Section 7.3), we began to descriptively explore whether *RtR* differentially affected sexual behaviors for youth who were already sexually experienced at the start of the study and youth who were not sexually experienced at baseline, but became sexually active throughout the course of the study.

In Exhibit D.1 below, we plotted the key behavioral outcomes (i.e., recent sexual activity and recent sexual intercourse without birth control[19]) for the treatment and control groups at each of the study data collection time points for all study youth (plotting symbol = A), youth who were never sexually active (plotting symbol = 0), youth who became sexually active during the study (plotting symbol = 1), and youth who were sexually active at baseline (plotting symbol = 2). In both graphs, we see notable differences in the trajectories of sexual behavior for each of the three groups, as would be expected. However, there were no discernible treatment-control differences between the group of youth who were sexually active when the study began (group 2) and those who initiated sexual behavior during the study period (group 1).

Exhibit D.1: Plots of Key Behavioral Outcomes Over Time for Subgroups of Youth Based on Sexual Activity Throughout the Study



Notes: For sexually active in last 90 days outcome, treatment/control sample sizes are as follows: Group 0: Baseline – 640/549, 12 months - 585/499, 24 months – 612/529; Group 1: Baseline – 469/326, 12 months – 444/305, 24 months – 465/323; Group 2: Baseline – 539/397, 12 months – 481/351, 24 months – 490/359. For sexual intercourse without contraception in last 90 days, treatment/control sample sizes are as follows: Group 0: Baseline – 640/549, 12 months – 585/499, 24 months – 612/529; Group 1: Baseline – 469/326, 12 months – 444/307, 24 months – 466/323; Group 2: Baseline – 538/399, 12 months – 481/351, 24 months – 490/359.

Wave 2 = short-term follow-up (12 months after baseline); Wave 3 = longer-term follow-up (24 months after baseline).

[19] Pregnancy is not included, because this outcome was only examined at the longer-term follow-up.

In further exploration, we made plots like the two shown above for additional sexual behavior outcomes for the full sample as well as for each of the study sites (not shown). These descriptive plots depicted very few consistent differences in treatment effects between the subgroups of youth who were sexually active at baseline, and those that became sexually active during the study period. Further analyses of differences between these groups would be exploratory and non-experimental, and thus are not included in this report.

Appendix E: Supporting Tables

Exhibit E.1: Characteristics of the Short-Term Follow-Up 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	.204
Grade				
Mean	9.25	9.27	-0.02	.626
Gender				
Female[c]	49.10	49.10	0.00	n/a
Race/ethnicity[d]				
Hispanic	46.09	47.12	-1.04	.518
Black	33.10	32.96	0.14	.903
White	11.34	10.73	0.61	.616
Other	9.54	9.18	0.36	.792
Family structure and relationships				
Lives with biological parents	93.06	92.24	0.82	.478
Feels very close to and cared for by father	45.46	46.74	-1.27	.564
Feels very close to and cared for by mother	63.38	65.98	-2.61	.171
Risk behaviors				
Ever smoked cigarettes	21.01	20.63	0.38	.826
Ever drank alcohol	45.73	45.05	0.69	.743
Ever used marijuana	31.23	30.00	1.23	.521
Sexual activity				
Ever sexually active[e]	30.57	31.32	-0.75	.683
Recently sexually active (in last 90 days)[e]	18.37	20.79	-2.42	.175
Sexual intercourse in the last 90 days	16.53	17.99	-1.46	.409
Oral sex in the last 90 days	12.12	14.56	-2.44	.114
Sexual risk behavior				
Sexual intercourse without a condom in the last 90 days	8.04	9.85	-1.80	.189
Oral sex without a condom in the last 90 days	10.40	11.77	-1.37	.348
Sexual intercourse without birth control in the last 90 days	5.36	6.71	-1.35	.254
Knowledge				
Knowledge of pregnancy risk[f]	51.84	50.61	1.22	.452
Knowledge of STI risk[f]	44.42	43.46	0.96	.411
Attitudes				
Attitudes toward protection[g]	3.04	3.04	0.00	.907
Intentions				
Intentions to have sexual intercourse[h]	41.14	39.16	1.98	.312
Intentions to have oral sex[h]	30.00	30.09	-0.08	.965

Measure	Treatment Mean[a]	Control Mean	Group Difference[b]	*p*-Value
Intentions to use birth control if they were to have sexual intercourse[h]	89.41	90.79	-1.38	.259
Intentions to use a condom if they were to have sexual intercourse[h]	94.59	94.17	0.42	.650

Source: Baseline survey administered prior to randomization.

Notes: Results in this table are based on the analytic sample of 2,368 - 2,689 short-term survey respondents who provided valid survey responses to relevant items on the baseline survey. 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 randomization 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] 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.

[f] Scores represent the average percentage of items answered correctly.

[g] Scale score averages responses ranging from 1 to 4. Higher scores indicate more positive attitudes.

[h] Intention to engage in the behavior in the next 12 months. Dichotomous variables, reported as percentage of respondents who responded affirmatively.

* $p<.05$, ** $p<.01$, *** $p< .001$ (two-tailed tests).

Exhibit E.2: Characteristics of the Longer-Term Follow-Up Analytic Sample at Baseline

Measure	Treatment Mean[a]	Control Mean	Group Difference[b]	*p*-Value
Demographic characteristics				
Age				
Mean	14.45	14.51	-0.05	.272
Grade				
Mean	9.23	9.24	-0.01	.774
Gender				
Female[c]	49.22	49.22	0.00	n/a
Race/ethnicity[d]				
Hispanic	49.34	50.37	-1.02	.556
Black	29.26	29.31	-0.05	.970
White	12.04	11.27	0.78	.547
Other	9.42	9.06	0.36	.803
Family structure and relationships				
Lives with biological parents	92.76	92.43	0.33	.776
Feels very close to and cared for by father	46.80	46.23	0.57	.791
Feels very close to and cared for by mother	63.54	66.07	-2.53	.178
Risk behaviors				
Ever smoked cigarettes	21.35	20.90	0.44	.806
Ever drank alcohol	46.15	45.24	0.91	.681
Ever used marijuana	31.03	29.26	1.76	.395
Sexual activity				
Ever sexually active[e]	30.53	30.67	-0.14	.934
Recently sexually active (in last 90 days) [e]	18.23	19.81	-1.58	.337
Sexual intercourse in the last 90 days	16.49	17.05	-0.56	.723
Oral sex in the last 90 days	11.72	14.15	-2.43	.092
Sexual risk behavior				
Sexual intercourse without a condom in the last 90 days	7.90	9.34	-1.44	.270
Oral sex without a condom in the last 90 days	9.95	11.23	-1.28	.364
Sexual intercourse without birth control in the last 90 days	4.90	6.51	-1.62	.163
Consequences of sexual risk behavior				
Ever pregnant or gotten someone pregnant (lifetime)	3.30	2.97	0.34	.633
Diagnosed with STI in the last 12 months	0.38	0.68	-0.29	.336
Knowledge				
Knowledge of pregnancy risk[f]	52.04	51.29	0.75	.672
Knowledge of STI risk[f]	44.17	43.56	0.60	.643
Attitudes				
Attitudes toward protection[g]	3.04	3.04	0.01	.714
Intentions				
Intentions to have sexual intercourse[h]	40.02	38.21	1.81	.334
Intentions to have oral sex[h]	29.58	30.40	-0.82	.648
Intentions to use a condom if they were to have sexual intercourse[h]	94.45	93.75	0.70	.448
Intentions to use birth control if they were to have sexual intercourse[h]	89.37	90.93	-1.57	.190

Source: Baseline survey administered prior to randomization.

Notes: Results in this table are based on the analytic sample of 2,452 – 2,799 longer-term survey respondents who provided valid survey responses to relevant items on the baseline survey. Values shown are percentages unless otherwise indicated. The items that compose measures of attitudes toward risky sexual 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 randomization 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] Sexual activity is defined differently across grantees. In Better Family Life, sexual activity refers to sexual intercourse, oral sex, and/or anal sex. In LifeWorks and San Diego Youth Services, students were not asked about anal sex.

[f] Scores represent the average percentage of items answered correctly.

[g] Scale score averages responses ranging from 1 to 4. Higher scores indicate more positive attitudes.

[h] Intention to engage in the behavior in the next 12 months. Dichotomous variables, reported as percentage of respondents who responded affirmatively.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests).

Appendix F: Supplemental Analyses on Unfavorable Site and Subgroup Findings

This appendix describes additional analyses we conducted to better understand the unfavorable effects of *RtR* on sexual behavior for youth in the LifeWorks site and for Hispanic youth. As discussed in Chapter 7, exploratory analyses revealed that in LifeWorks, but not the other two sites, students enrolled in *RtR* were more likely to be recently sexually active, to have had sexual intercourse in the last 90 days, and to have had oral sex in the last 90 days, compared with students in the control group. Similarly, Hispanic students who were enrolled in *RtR* were more likely to be recently sexually active, and more specifically to have engaged in oral sex in the last 90 days, compared to their control group counterparts. Given that a majority of youth in LifeWorks were Hispanic, it was plausible that these findings reflected the same subgroup. To disentangle these unfavorable site and subgroup effects, we conducted additional analyses.

First, we tested the three-way interaction between treatment status*site[20]*Hispanic predicting sexual activity[21] in order to test whether the LifeWorks site effect also depended on Hispanic ethnicity or vice versa. The three-way interaction was not significant (p = .606). Therefore the difference in effect between Hispanic and non-Hispanic youth (indicating an unfavorable effect among Hispanic youth) was not dependent on site, and the difference in effects between sites (indicating an unfavorable effect in LifeWorks) was not dependent on racial/ethnic subgroup. In other words, it was not that *RtR* only had an unfavorable effect on Hispanic youth in LifeWorks.

Next, we ran the impact model including both the treatment*site[22] and treatment*Hispanic 2-way interactions in the same model. Both 2-way interactions were significant (p = .031 and p = .049, respectively) indicating that when controlling for the site-level difference, there was still a difference in effect between Hispanic and non-Hispanic youth, and when controlling for the difference between Hispanic and non-Hispanic youth, there was still a difference between sites. In other words, the unfavorable effect in LifeWorks was not explained by the unfavorable effect among Hispanic youth, and the unfavorable effect among Hispanic youth was not explained by an unfavorable effect in LifeWorks. We conclude that there were unique unfavorable program effects both for Hispanic youth and for youth in the LifeWorks site; neither effect drove the other.

Aside from Hispanic ethnicity, another characteristic that distinguished LifeWorks from the other sites was the age of the youth. Youth in LifeWorks were, on average, 15.2 years old at baseline compared with 14.7 in BFL and 13.7 in SDYS. Accordingly, we further explored whether age might have explained the unfavorable effects found in LifeWorks. When looking within the LifeWorks sample, we found no significant differences in effects between older and younger youth on any of the behavioral outcomes. However, when looking within the sample of youth aged 15 or older, we did find a significant difference between sites such that there was an unfavorable and significant effect in LifeWorks. In the other two sites, the effects were not significant, but they were in the favorable direction (see Exhibit F.1). Given that

---

[20] BFL was excluded from this analysis, because only 3 percent of the BFL sample was Hispanic.

[21] Here we present analyses conducted using sexual activity as the outcome. Other behavioral outcomes yielded similar findings.

[22] BFL was excluded from this analysis, because only 3 percent of the BFL sample was Hispanic.

the difference between LifeWorks and the other two sites persisted even when looking specifically at the sample of older youth, we concluded that the unfavorable effects in LifeWorks were not driven by age.

Exhibit F.1: Differences in Effects on Sexual Activity by Site, Among Youth Aged 15+

Outcome	Better Family Life			LifeWorks			San Diego Youth Services			*p*-Value for the Test of Differences Across Sites[a]
	n	T Effect[b]	*p*-Value	n	T Effect[b]	*p*-Value	n	T Effect[b]	*p*-Value	
Recently sexually active	439	-8.26	0.08	605	10.00*	0.01	66	-11.40	0.38	.008**
Sexual intercourse in the last 90 days	440	-7.52	0.12	605	10.81*	0.01	66	-6.44	0.62	.014*
Oral sex in the last 90 days	439	-7.76	0.08	602	7.10	0.07	66	-10.45	0.40	.029*

Source: Follow-up survey administered 24 months after baseline.

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

[b] The treatment effect was estimated in a regression model that controls for randomization blocks and other covariates. The treatment effect is expressed as a difference in percentage points.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests).