



¡Cúdate!:
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

Cristofer Price

Michelle Blocklin

Abt Associates

10 Fawcett Street
Cambridge, MA 02138

In Association with:

**Decision Information
Resources, Inc.**

and

Belmont Research Associates

Contents

1. Introduction.....1

1.1 The TPP Replication Study1

1.2 The Three Models Replicated2

1.3 Focus of This Report.....2

2. The Program Model: ¡Cuidate!3

2.1 ¡Cuidate! Logic Model4

2.2 Prior Evidence of Effectiveness5

3. Evaluation Design and Implementation.....7

3.1 Research Questions7

3.2 Key Design Features8

3.3 Measures and Measurement Schedule8

3.4 Analytic Approach10

3.5 Implementing the Study Design.....11

3.6 Conducting the Analysis17

4. Implementation Findings and Baseline Characteristics.....21

4.1 Program Implementation.....21

4.2 Sample Characteristics23

5. Program Impacts on Youth Sexual Activity, Sexual Risk Behavior, and Consequences of Sexual Risk Behavior.....26

5.1 Confirmatory Analyses of Program Impacts on Key Behavioral Outcomes26

5.2 Exploratory Analyses of Effects on Additional Behavioral Outcomes.....27

6. Exploratory Analyses of Program Effects on Non-Behavioral Intermediate Outcomes.30

6.1 Knowledge of Pregnancy and STI Risk30

6.2 Attitudes31

6.3 Motivation to Delay Childbearing32

6.4 Intentions.....32

6.5 Skills.....33

7. Exploratory Analyses of Program Effects by Site and Subgroup35

7.1 Site-Level Differences35

7.2 Subgroup Differences.....36

7.3 Differences in Program Effect on Youth Based on Sexual Activity Throughout the Study38

8. Discussion40

References.....42

Appendix A: Measures44

Appendix B: Site-Level Effects.....52

Appendix C: Subgroup Effects.....57

Appendix D: Supplemental Analyses on the Effect on Pregnancy61

Appendix E: Sensitivity Analyses Conducted on Hispanic-Only Sample69

**Appendix F: Preliminary Descriptive Findings on Subgroups of Youth Based on
Sexual Activity Throughout the Study80**

Appendix G: Supporting Tables.....81

Exhibits

Exhibit 2.1: *¡Cuidate!* Modules, Topics, and Core Element(s) 3

Exhibit 2.2: Logic Model for *¡Cuidate!*..... 5

Exhibit 3.1: Outcome Measures 8

Exhibit 3.2: Measurement Domains and Key Outcomes..... 11

Exhibit 3.3: Treatment and Control Conditions in the Three Replication Sites 13

Exhibit 3.4: Study Samples..... 16

Exhibit 3.5: *¡Cuidate!* Survey Response Rates 17

Exhibit 4.1: Baseline Characteristics of the Longer-Term Analytic Sample by Site 24

Exhibit 5.1: Short-Term and Longer-Term Impacts on Key Behavioral Outcomes..... 27

Exhibit 5.2: Additional Short-Term and Longer-Term Effects on Sexual Activity, Sexual Risk Behavior, and Consequences 28

Exhibit 6.1: Short-Term and Longer-Term Effects of *¡Cuidate!* on Knowledge 31

Exhibit 6.2: Short-Term and Longer-Term Effects of *¡Cuidate!* on Attitudes..... 31

Exhibit 6.3: Short-Term and Longer-Term Effects of *¡Cuidate!* on Motivation to Delay Childbearing..... 32

Exhibit 6.4: Short-Term and Longer-Term Effects of *¡Cuidate!* on Intentions 33

Exhibit 6.5: Short-Term and Longer-Term Effects of *¡Cuidate!* on Skills 33

Exhibit A.1: Youth Sexual Activity, Sexual Risk Behavior, and Sexual Consequences Measures 44

Exhibit A.2: Knowledge Scales and Component Items..... 47

Exhibit A.3: Attitudes Scales and Component Items 48

Exhibit A.4: Motivation Scale and Component Items 50

Exhibit A.5: Intentions Measures 50

Exhibit A.6: Skills Scales and Component Items 51

Exhibit B.1 Short-Term Effects on Sexual Activity and Sexual Risk Behavior by Site 53

Exhibit B.2: Longer-Term Effects on Sexual Behavior, Sexual Risk Behavior, and Consequences by Site 54

Exhibit B.3: Short-Term Effects on Non-Behavioral Intermediate Outcomes by Site 55

Exhibit B.4: Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Site 56

Exhibit C.1: Short-Term Effects on Behavioral Outcomes by Subgroup 57

Exhibit C.2: Longer-Term Effects on Behavioral Outcomes by Subgroup 58

Exhibit C.3: Short-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup 59

Exhibit C.4: Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup 60

Exhibit D.1: Differences in Effect on Pregnancy by Age, within Sites..... 62

Exhibit D.2: Hispanic/Non-Hispanic Differences in Effect on Pregnancy, within Sites..... 62

Exhibit D.3: Characteristics of the Longer-Term Analytic Sample in La Alianza at Baseline 63

Exhibit D.4: A Simple Mediation Model 65

Exhibit D.5: Summary of Results of Multiple Mediation Model 68

Exhibit E.1:	Short-Term and Longer-Term Effects on Sexual Activity, Sexual Risk Behavior, and Consequences for Hispanic Sample Only.....	69
Exhibit E.2:	Short-Term Effects on Non-Behavioral Intermediate Outcomes for Hispanic Sample Only	71
Exhibit E.3:	Longer-Term Effects on Non-Behavioral Intermediate Outcomes for Hispanic Sample Only	72
Exhibit E.4:	Short-Term Effects on Sexual Activity and Sexual Risk Behavior by Site for Hispanic Sample Only	73
Exhibit E.5:	Longer-Term Effects on Sexual Behavior, Sexual Risk Behavior, and Consequences by Site for Hispanic Sample Only	74
Exhibit E.6:	Short-Term Effects on Non-Behavioral Intermediate Outcomes by Site for Hispanic Sample Only	75
Exhibit E.7:	Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Site for Hispanic Sample Only	76
Exhibit E.8:	Short-Term Effects on Sexual Behavior and Sexual Risk Behavior by Subgroup for Hispanic Sample Only	77
Exhibit E.9:	Longer-Term Effects on Sexual Behavior, Sexual Risk Behavior, and Sexual Consequences by Subgroup for Hispanic Sample Only	77
Exhibit E.10:	Short-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup for Hispanic Sample Only	78
Exhibit E.11:	Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup for Hispanic Sample Only	79
Exhibit F.1:	Plots of Confirmatory Outcomes Over Time for Subgroups of Youth Based on Sexual Activity Throughout the Study	80
Exhibit G.1:	Characteristics of the Short-Term Follow-Up Analytic Sample at Baseline	81
Exhibit G.2:	Characteristics of the Longer-Term Follow-Up Analytic Sample at Baseline	83

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 a 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 Study¹ was conducted for HHS, under contract with the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and OAH, by Abt Associates and its subcontractors, Belmont Research Associates and 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
- *¡Cuidate!*, an HIV/STI risk reduction program targeting Latino youth.

Criteria used in the selection of these models included the breadth and scale of the proposed replication effort; and the number of grantees that proposed to replicate a model. 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 *¡Cuidate!*, 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: ¡Cuidate!

¡Cuidate! is an HIV/STI risk reduction curriculum adapted from the *Be Proud! Be Responsible!* curriculum and culturally tailored for Latino youth. It aims to reduce the risk of STIs, in particular HIV, by affecting sexual behaviors such as sexual intercourse, number of sexual partners, and condom use. Six 60-minute modules are delivered in small groups of six to 10 youth, led by a trained adult facilitator who is bilingual in English and Spanish, although the program is delivered in English only.

¡Cuidate! was originally tested in an after-school setting on consecutive weekends, but it can be delivered in other settings and on different schedules (Villaruel, Jemmott, & Jemmott, 2005). The curriculum modules are delivered in participatory, interactive sessions. Each session weaves in the theme of “taking care”—of oneself, one’s partner, family, and community. Exhibit 2.1 shows the topics covered in each of the six modules and links them to the program’s core elements.

Exhibit 2.1: ¡Cuidate! Modules, Topics, and Core Element(s)

Module	Topic/Activities	Core Elements Addressed ^a
Introduction and Overview	<ul style="list-style-type: none"> Getting to Know You Talking Circle Creating Group Rules HIV/AIDS What It Means to Be Latino/Latina Cultural Values What Latinos Think About HIV/AIDS and Safer Sex 	1, 2, 3, 6
Building HIV Knowledge	<ul style="list-style-type: none"> View <i>¡Cuidate!</i> Video Myths and Facts 	1, 2, 3, 6
Understanding Vulnerability to HIV Infection	<ul style="list-style-type: none"> Acknowledging the Threat of HIV/AIDS Latino Cultural Values and HIV “A Romance” (role play) La Lotería Talking Circle 	1, 2, 3, 5, 6
Attitudes and Beliefs about HIV/AIDS	<ul style="list-style-type: none"> Welcome and Talking Circle Music and Discussion Quién es más macho? Quién es más mujer? Adolescent Vulnerability to HIV La Zona Peligrosa 	1, 2, 3, 6
Building Condom Skills	<ul style="list-style-type: none"> Discussing Condoms Condom-Use Skills Overcoming Barriers to Condom Use What Gets in the Way of Caring Behavior? Condom Line-Up 	1, 3, 4, 5, 6
Building Negotiation and Refusal Skills	<ul style="list-style-type: none"> No Hay Razon How to Use the S.W.A.T. Technique and Scripted Role Plays AIDS Jeopardy Game Talking Circle 	1, 2, 3, 5, 6

Source: *¡Cuidate!* Starter Kit

^a Core elements: (1) Incorporating the theme of “taking care” of oneself and one’s partner, family, and community throughout the program. (2) Using culturally and linguistically appropriate materials and activities to show and emphasize core Latino cultural values, specifically familialism and gender roles and how those are consistent with safer sex behavior. (3) Incorporating activities that increase knowledge and influence positive attitudes, beliefs, and self-efficacy regarding HIV sexual risk-reduction behaviors. (4) Modeling and practicing the effective use of condoms. (5) Building participants’ skills in problem solving, negotiation of safe sex, and refusal of unsafe sex. (6) Delivering sessions in highly participatory, interactive small groups.

2.1 ¡Cuidate! Logic Model

The materials used in the ¡Cuidate! sessions emphasize core Hispanic values and feelings, linking them to safer sexual behavior. The facilitator demonstrates correct condom use and teaches negotiation and refusal skills. Youth are exposed to information about HIV/STI transmission and prevention. Through active participation in discussions, sharing ideas and feelings, and role-playing situations in which they may be pressured to have unwanted or unsafe sex, participants are exposed to more information to increase their knowledge and understanding of how to avoid unsafe sexual behaviors; improve their attitudes, values, and beliefs; and increase their motivation to delay childbearing. These discussions, together with repeated role-play activities, strengthen intentions to abstain from sexual activity and to use protection. Role plays also support the acquisition of skills youth need to deal with unwanted pressures and risky situations, refuse unsafe sexual behavior and negotiate safer sexual behavior, and use condoms correctly.

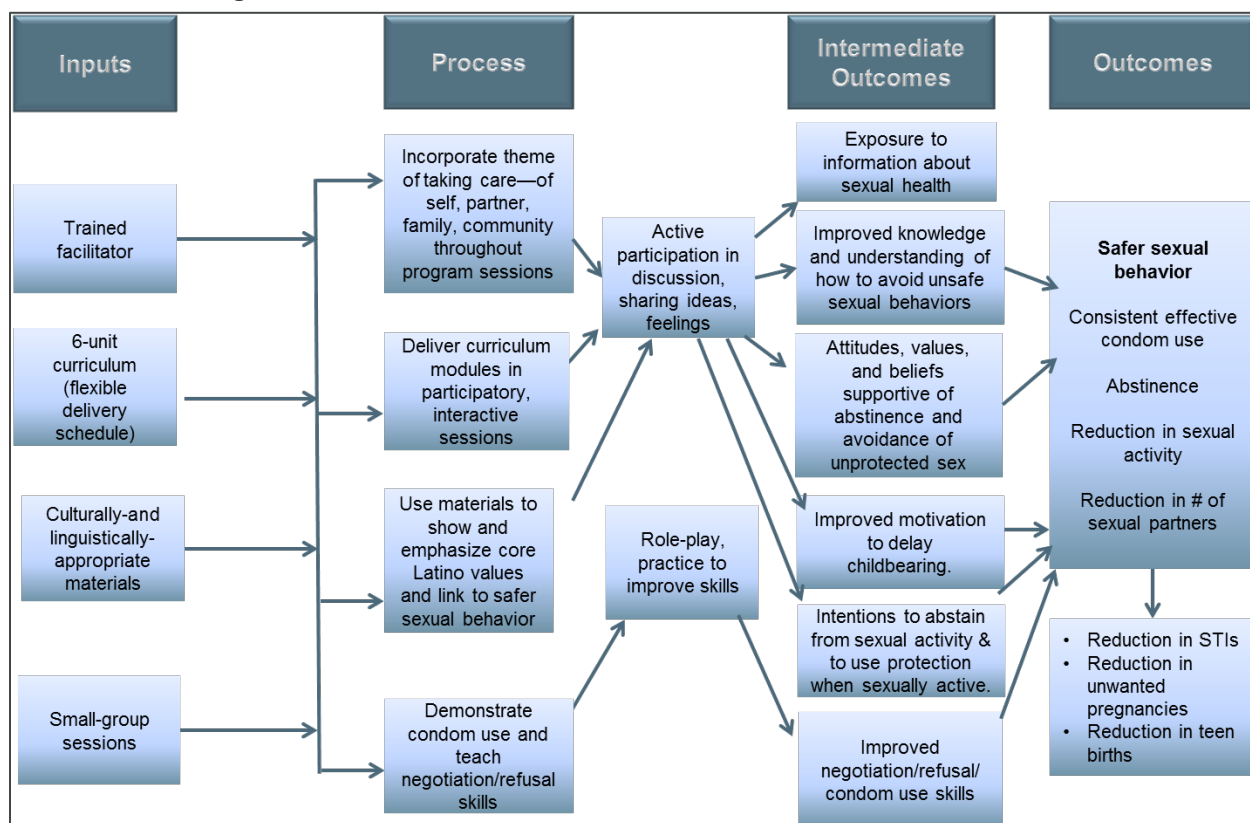
These non-behavioral outcomes (exposure to information; positive changes in knowledge, attitudes, motivation, intentions, and skills) are hypothesized to lead to the safer behavioral outcomes that the program seeks to achieve: correct and consistent use of condoms, abstinence from sex, reduced sexual activity, and reduced number of partners. Safer sexual behavior is ultimately expected to result in a reduction in the rates of STIs among teens, as well as a reduction in pregnancy rates and births.⁴

Exhibit 2.2 below shows the program elements, the hypothesized outcomes, and the pathways by which the program seeks to achieve these outcomes. The model implies but does not specify time periods between early non-behavioral, “intermediate” outcomes and later behavioral outcomes they 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., at ages 13-14), ¡Cuidate!, like many other TPP 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. 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., ages 17-19), ¡Cuidate! 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, over a much shorter period of time, that youth are already engaged in so those behaviors are less risky. If the program is able to effect 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.

⁴ In 2012, the curriculum was revised to include material on pregnancy prevention. However, the grantees whose projects were funded in 2010 were trained on and implemented the original curriculum, supplementing it with additional sessions on pregnancy prevention or incorporating that topic into existing sessions.

Exhibit 2.2: Logic Model for ¡Cuidate!



Though the general guidance is that ¡Cuidate! 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 ¡Cuidate! program model proposed by each of the organizations replicating it. They made modifications to comply with state mandates, to address gaps in program content, or to accommodate local constraints, though still adhering to the core elements of the model.

2.2 Prior Evidence of Effectiveness

¡Cuidate! is one of several programs that addresses the issue of sexual risk behavior among Hispanic adolescents, and one of the few that met the standards for having evidence of effectiveness, and therefore was eligible for replication funding under the TPP Program (HHS, 2010). Aside from the single research study of ¡Cuidate! cited in the evidence review (Villaruel, Jemmott, & Jemmott, 2006), few studies of pregnancy or HIV/STI prevention have focused on Hispanic youth, although a few have included substantial numbers of Hispanic adolescents as part of a larger population.

¡Cuidate! was developed more than a decade ago in response to several concerns. Hispanics, and Hispanic youth in particular, constitute a large and growing portion of the U.S. population. Since 1995,

Hispanic children have been the largest minority group of children in the United States. By 2011, 23 percent of the population under the age of 18 was Hispanic (U.S. Census Bureau, 2012). This population is disproportionately affected by HIV/AIDS. In 2010, the rate of HIV/AIDS in Hispanic adults/adolescents was more than three times the rate in the White population, and Hispanics accounted for 21 percent of all new HIV infections (CDC, 2012). This disproportion in rates has remained stable since 2001. Among possible reasons for the disparity are that Hispanic adolescents are less likely to use condoms and more likely to report multiple sexual partners compared to other racial/ethnic groups (CDC, 2004). In 2011, only 54.9 percent of Hispanic youths reported using a condom during sex, compared with 63.3 percent and 62.4 percent of White and African-American youths, respectively (American Academy of Pediatrics, 2013).

Be Proud, Be Responsible!, the HIV reduction program on which *¡Cuidate!* is based, was shown to be effective for African-American youth; it also was one of the 28 program models that had evidence of effectiveness (Goesling et al., 2014; Jemmott, 1992; Jemmott et al., 1999). *¡Cuidate!* was tested by the developer among mostly Puerto Rican youth in Philadelphia. Youth in the program reported fewer incidents of sexual intercourse, fewer sexual partners, and fewer days of unprotected intercourse than did youth assigned to a health promotion program (Villaruel et al., 2006). In an earlier paper (Villaruel et al., 2005), the developer notes the diversity among Hispanics in the United States and its implications for adolescent risk behavior (HIV/AIDS rates are highest among Puerto Ricans in the Northeast and lowest among Mexican-Americans in the West), but does not restrict the generalizability of the study's findings to the Hispanic subgroup on which it was tested. Because two of the three replications in the TPP Replication Study were conducted in Western states, with primarily Mexican-American populations, this evaluation sheds some light on the generalizability of the effectiveness of the intervention.

3. Evaluation Design and Implementation

The impact study is designed to estimate the effects of *¡Cuídate!* 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 *¡Cuídate!* 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 *¡Cuídate!* have an impact on sexual behavior after six months and 18 months?
2. Did *¡Cuídate!* reduce the incidence of unplanned pregnancies after 18 months?
3. Did *¡Cuídate!* reduce the incidence of STIs after 18 months?
4. Did *¡Cuídate!* have an impact on non-behavioral, intermediate outcomes hypothesized to lead to behavior change (i.e., knowledge, attitudes, motivation, intentions, and skills) after six months and 18 months?
5. Do program impacts 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 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 impact for specific sites and subgroups.

We elected to investigate non-behavioral and 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 would probably 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 impacts on so many outcomes does, however, pose challenges for the ways in which data are analyzed and how the results are interpreted. The sheer number of statistical tests of

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.

effectiveness means that we would expect five percent to generate statistically significant results simply by chance. In a later section of this chapter, we describe the steps we took to minimize the risk of incorrectly concluding that *¡Cuidate!* had an impact.

3.2 Key Design Features

The design of the evaluation of *¡Cuidate!* 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 individual youths were randomly assigned either to a group that received the *¡Cuidate!* intervention or to a control group that did not;
- 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; and
- A strategy that identifies a key set of 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	
<ul style="list-style-type: none"> • Recent sexual activity (in last 90 days)^a • Sexual intercourse in the last 90 days • Oral sex in the last 90 days • Anal sex in the last 90 days 	Single items, scored 1 (yes) or 0 (no).
<ul style="list-style-type: none"> • 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, scored 1 (yes) or 0 (no).
Sexual risk behavior	
<ul style="list-style-type: none"> • 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) • Anal sex without a condom (in last 90 days) 	Single items, scored 1 (yes) or 0 (no).

EVALUATION DESIGN AND IMPLEMENTATION

Measure	Definition
Sexual Consequences (longer-term follow-up only)	
<ul style="list-style-type: none"> • Pregnant or gotten someone pregnant since baseline^a • Diagnosed with STI in the last 12 months 	Single items, scored 1 (yes) or 0 (no).
Non-Behavioral Intermediate Outcomes	
Knowledge	
<ul style="list-style-type: none"> • Knowledge of pregnancy risk 	<u>Continuous index:</u> Average of responses to four questions about circumstances in which it is possible to become pregnant and the extent to which contraceptive methods protect 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.
<ul style="list-style-type: none"> • Knowledge of STI risk 	<u>Continuous index:</u> 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	
<ul style="list-style-type: none"> • Attitudes toward protection 	<u>Continuous index:</u> 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.
<ul style="list-style-type: none"> • Attitudes toward risky sexual behavior 	<u>Continuous index:</u> 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	
<ul style="list-style-type: none"> • Motivation to delay childbearing 	<u>Continuous index:</u> 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 the next 12 months)	
<ul style="list-style-type: none"> • 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 item, scored 0 or 1, with 1 representing stronger intention.
Skills	
<ul style="list-style-type: none"> • Refusal skills 	<u>Continuous index:</u> 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.
<ul style="list-style-type: none"> • Condom negotiation skills 	<u>Continuous index:</u> 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); six months after the baseline survey (short-term follow-up); and 18 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. The schedule was also similar to the measurement schedule of the original study (Villaruel et al., 2005).

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 *¡Cuidate!*, 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 the replication site (e.g., in one replication site, at the grantee's request, surveys of middle school 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 *¡Cuidate!* 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,⁷ and data collection procedures described elsewhere in this chapter 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.

⁷ With the minor exception of the exclusion of questions about anal sex in the one site with a notably younger population.

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 outcomes of greatest interest to federal policymakers.

The first step was to identify a small set of behavioral outcomes by which the success of *¡Cuidate!* 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 *¡Cuidate!* 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 effectiveness of *¡Cuidate!* are referred to as “confirmatory analyses.” Based on the decisions described above, our confirmatory analyses estimate the impacts of *¡Cuidate!* on the **key outcomes for the full sample**, using data pooled across the three replication sites. Additional analyses, testing different outcomes or different samples or subgroups are referred to as “exploratory analyses” and should be interpreted as suggestive of potential program effects (see Schochet, 2008b).

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. In each of the selected replication sites, the services provided to youth in the intervention group had to be substantially different from the services provided to youth in the control group. In addition, grantees needed to be able to recruit enough youth over two years to meet the needs of the study. Selection was constrained because the

¡Cuidate! intervention was being replicated in only four TPP Program sites, and these replications were not planned with the requirements of a rigorous evaluation in mind.⁸ It was apparent that one of the four sites would not be able to build the sample of youth needed for the study over a period of two years. Combined with other considerations that could impede a strong test of the model, this led us to eliminate one of the four potential candidates.

The three grantees selected are described below.

- **Community Action Partnership of San Luis Obispo County.** Community Action Partnership, a non-profit agency founded in 1965 and based in San Luis Obispo, California, provides a wide variety of programs and services to residents of San Luis Obispo County and 10 other California counties. Since 1977, the agency has provided comprehensive sexual health education programming in schools for youth ages 10–18. The agency also has its own reproductive health clinics, including teen-designed and peer-provided teen clinics.
- **La Alianza Española.** Founded in 1970 and based in Boston, Massachusetts, La Alianza is a non-profit advocacy and service organization whose core programs address family mental health, public health, and workforce education. The agency has worked with the Boston Housing Authority to provide information about HIV/STI and pregnancy prevention to young Latina women and, with other members of the Adolescent Trials Network (a collaborative of community-based organizations and health care providers, based at Boston Children’s Hospital), to reduce HIV infection rates among adolescents.
- **Touchstone Health Services.**⁹ This non-profit organization has more than 30 years of experience providing behavioral and mental health prevention and treatment programs and services to youth across the Greater Phoenix, Arizona, area. The agency has focused its prevention work on the Maryvale community, which has a predominantly Hispanic population. Before receiving the TPP grant, Touchstone had implemented a substance abuse prevention program and some limited sexual health programming in schools in this community.

Of the three selected grantees, only La Alianza had some limited prior experience with *¡Cuidate!* In 2009, the agency implemented a few program cycles in a small number of Boston public schools.

3.5.2 Site-Specific Program Designs

The three replications of *¡Cuidate!* 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 2.2. Fidelity was assessed, monitored, and reported to OAH at regular intervals by program staff. Beyond these core elements, there were small variations in program design across the three sites. The

⁸ The 2010 TPP grant program included 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 ranges (greater than \$1 million per year) were expected to be implemented in multiple sites within a targeted geographic area and were required to have an independent local evaluation. The *¡Cuidate!* replications selected for the study were in the lower funding range (less than \$1 million per year) and so were not expected to have a rigorous local evaluation.

⁹ During the time of the study, the agency was called Touchstone Behavioral Health.

Community Action Partnership replication added two sessions to comply with the California requirements governing sex education—one on birth control and a second on STIs other than HIV. All three replications added a brief review of reproductive anatomy (to ensure that participants were aware of correct names for body parts). Touchstone also added a session on pregnancy prevention. With OAH’s approval, Touchstone also delivered the program in larger groups, of 20 to 25 youths, with two health educators rather than one.

3.5.3 Settings for the Program

¡Cuidate! was delivered in both school and community settings. In San Luis Obispo County, Community Action Partnership staff implemented the program in 10th-grade classrooms in three public high schools across the county. In Greater Boston, La Alianza selected settings for the program that were more diverse: three public high schools (one traditional, one vocational-technical, and one charter school) and two community-based organizations (as part of a summer youth employment program and a summer sports program). In two of those three schools, La Alianza staff implemented the program in classrooms during the regular school day (ninth- and 12th-grade classrooms in one school and mixed-grade classrooms in the other). In the third school, La Alianza staff implemented the program after regular school hours. In Phoenix, Touchstone staff implemented the program in eighth-grade classrooms in 11 K-8 schools.

3.5.4 Treatment and Control Conditions

Across the three replications, youths assigned to the treatment group were offered all of the sessions of *¡Cuidate!*, delivered by health educators who were members of the grantee agency staff. Health educators were trained by the program developer and then supervised and monitored continuously by grantee supervisory staff 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 those important commonalities, there were differences in the treatment and control conditions across and within replication sites, as shown in Exhibit 3.3 below. The treatment condition varied across all three replication sites in total number of sessions delivered, schedule for delivery, and group gender and size. As for the control condition, in two of the three replication sites, students participated in a standardized activity (i.e., regularly scheduled physical education classes in Community Action Partnership; a healthy lifestyle curriculum in Touchstone). In the third replication site (La Alianza), where the program was delivered in a variety of settings, control group activities varied greatly.

Exhibit 3.3: Treatment and Control Conditions in the Three Replication Sites

Settings	Treatment Condition	Control Condition
Community Action Partnership of San Luis Obispo		
10th-grade physical education classes in three public high schools in San Luis Obispo County, CA	<p><u>Number of sessions:</u> 6 one-hour ¡Cuidate! sessions plus two additional sessions (one on birth control, the other on STIs other than HIV)</p> <p><u>Delivery schedule:</u> Weekly</p> <p><u>Gender and size of groups:</u> Small mixed-gender groups</p>	Regular weekly physical education classes

EVALUATION DESIGN AND IMPLEMENTATION

Settings	Treatment Condition	Control Condition
La Alianza Hispana		
<i>Multiple Settings in Greater Boston, MA:</i>		
Ninth- and 12th-grade physical education classes in a technical high school	<u>Number of sessions:</u> eight 45-minute <i>¡Cuidate!</i> sessions <u>Delivery schedule:</u> Varied; daily or weekly depending on the grade <u>Gender and size of groups:</u> Small mixed-gender groups	Regular physical education classes
Mixed-grade (ninth–12th) health classes in a public charter school	<u>Number of sessions:</u> 6 one-hour <i>¡Cuidate!</i> sessions <u>Delivery schedule:</u> Weekly <u>Gender and size of groups:</u> Small mixed-gender groups	Regular weekly health classes
After-school program in a public high school	<u>Number of sessions:</u> 6 one-hour <i>¡Cuidate!</i> sessions <u>Delivery schedule:</u> Daily <u>Gender and size of groups:</u> Small mixed-gender groups	No alternative provided for control group youth, but they were free to enroll in other after-school activities
Early evening sports program serving a mix of ages in a community agency	<u>Number of sessions:</u> 3 two-hour <i>¡Cuidate!</i> sessions <u>Delivery schedule:</u> Daily <u>Gender and size of groups:</u> Small mixed-gender groups	Control group youth were free to participate in sports activities offered
Summer youth employment program in community agency, during a period of the day when youth are free to choose from a variety of enrichment activities	<u>Number of sessions:</u> 2 three-hour <i>¡Cuidate!</i> sessions <u>Delivery schedule:</u> Weekly <u>Gender and size of groups:</u> Small mixed-gender groups	Control group youth were free to choose an activity (youth basketball was a frequent choice)
Touchstone Behavioral Health		
Eighth-grade classes in 11 public schools (K-8) in a single school district in Phoenix, AZ	<u>Number of sessions:</u> 6 one-hour <i>¡Cuidate!</i> sessions plus two additional sessions (one on pregnancy prevention, one to provide extra time for completion) <u>Delivery schedule:</u> Weekly <u>Gender and size of groups:</u> Single-gender groups of up to 25 youths, with two co-facilitators	8 one-hour sessions of a healthy lifestyle curriculum. Single-gender groups met with two co-facilitators three times a week over a three-week period. Topics covered self-esteem, body image, nutrition and exercise, stress management, decision-making skills, consequences of underage drinking, dangers of marijuana and other drugs, peer pressure

3.5.5 Recruitment and Random Assignment

Individual youths were recruited for the study within each of the replication sites. In each school that agreed to participate, school staff identified classes or time slots in which *¡Cuidate!* would be offered (small groups of students would be pulled out of regular classes for the program). Agency staff then recruited students from those classes. In La Alianza, where the program was implemented in summer youth employment and sports programs, youth were recruited directly from the participant list.

To recruit youths for the study, agency staff members who had been carefully trained by the Abt evaluation team conducted presentations to the identified groups or classes. These presentations included information about the study procedures, a practical illustration of random assignment, and a description of the treatment and control conditions. The presentations were intended to personalize the study and help with recruitment.

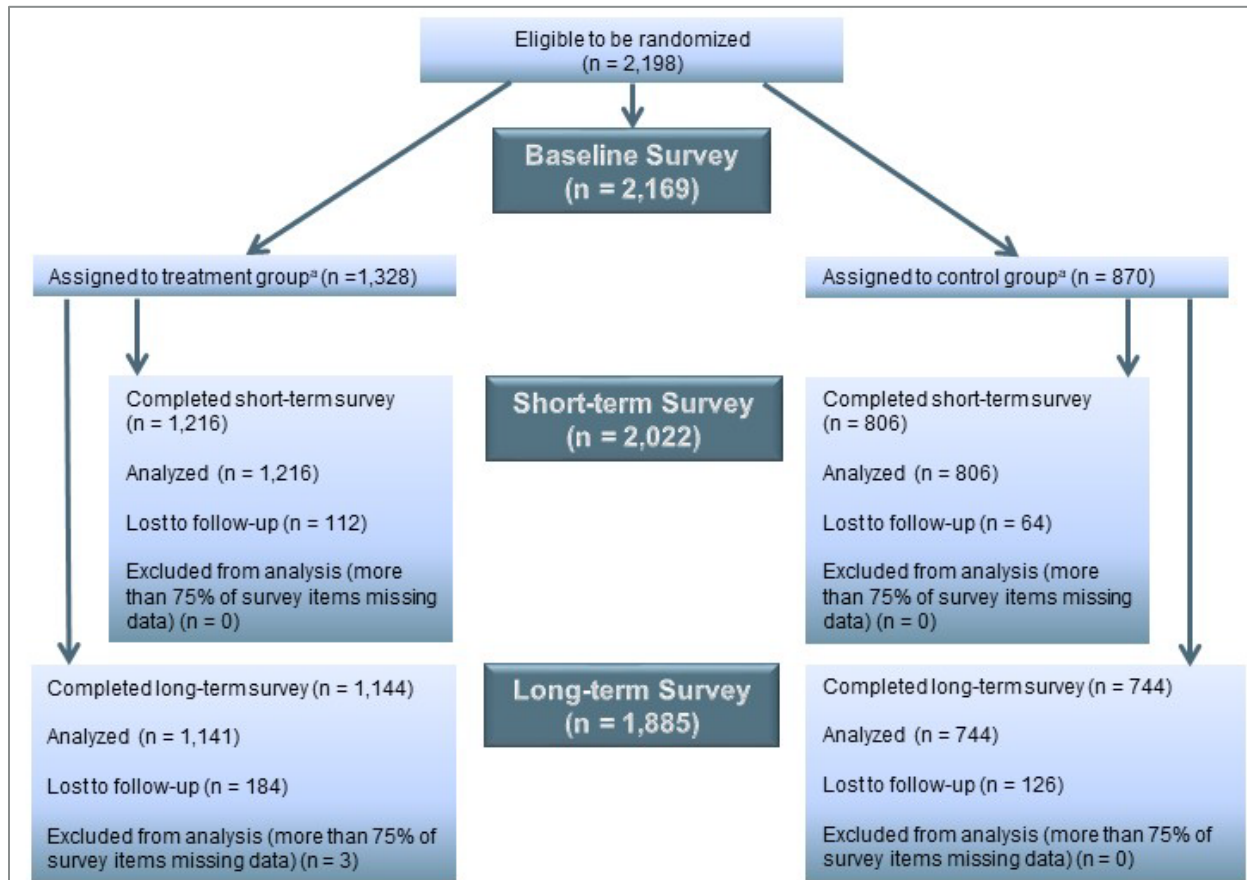
The same agency staff distributed parent consent forms and study brochures and provided teachers with small incentives (\$5 gift cards) to offer students for the return of parent consent forms. Agency staff worked with individual teachers to gather the consent forms (both refusals and agreements) and notified Abt study staff of students who had parent consent. Of consent forms returned, 89 percent granted youth permission to participate in the study.¹⁰

Those youth with permission were invited to complete the baseline survey and were included in the random assignment conducted by Abt. Youth with consent for participation were randomly assigned to either a treatment group that received *¡Cuidate!* or a control group that did not. In La Alianza and Community Action Partnership, the two sites that offered *¡Cuidate!* in small groups, individual students were randomly assigned within gender, class periods, semesters, and schools using a 2:1 ratio. That is, for every two youth assigned to the treatment group, one youth was assigned to the control group. In Touchstone, where the program was delivered in larger, single-gender groups, youth were randomly assigned within gender, semester, and school using a 1:1 ratio. Youth who did not return a consent form or who did not have parental consent were not included in the study and instead were assigned to the class with the control group. Participants were informed of their study assignment only after completing the baseline survey.

Exhibit 3.4 shows how we arrived at the study's analytic samples (for the short and longer term) via the random assignment and survey completion processes, beginning with 2,198 eligible youth (i.e., those who had parental permission to participate).

¹⁰ Data were not collected on youths whose parents declined consent. Therefore, it is not possible to assess similarities and differences between consented and non-consented youths.

Exhibit 3.4: Study Samples



^a A total of 29 participants did not take the baseline survey. Among them, 17 youth were assigned to the treatment group and 12 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 in 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, study participants completed the baseline survey in group sessions in schools or other settings, using school computers where possible or tablets dedicated to the study where it was not. Study staff oversaw the survey, distributing survey incentives (gift cards) to youths upon completion. Those who were absent on the day of the survey were contacted and a follow-up session was arranged. Of the 2,198 eligible youths, 2,169 (98.7 percent) completed a baseline survey.

For the two follow-up surveys, six months after baseline (short-term) and 18 months after baseline (longer-term), only the web-based ACASI system was used. Study participants could access the survey and complete it using their 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 upon completion.

As Exhibit 3.5 shows, 92 percent of eligible youths completed the short-term follow-up survey, and 86 percent of eligible youths completed the longer-term follow-up survey. At both data collection points, there was almost no difference in the response rates of youth in the treatment group compared with youth in the control group. Of the three replication sites, La Alianza had the highest response rate at the short-term and longer-term follow-ups.

Exhibit 3.5: ¡Cúidate! Survey Response Rates

	All Participants			Completed Short-Term Follow-Up						Completed Longer-Term Follow-Up					
				Total		Treatment		Control		Total		Treatment		Control	
	Total N	Treatment	Control	N	%	N	%	N	%	N	%	N	%	N	%
All Sites	2,198	1,328	870	2,022	92.0	1,216	91.6	806	92.6	1,885	85.8	1,141	85.9	744	85.5
Community Action Partnership	530	313	217	490	92.5	287	91.7	203	93.6	450	84.9	262	83.7	188	86.6
La Alianza	728	463	265	692	95.1	439	94.8	253	95.5	656	90.1	421	90.9	235	88.7
Touchstone	940	552	388	840	89.4	490	88.8	350	90.2	779	82.9	458	83.0	321	82.7

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 group versus control group members using a regression framework, in which we include baseline covariates to increase statistical precision (i.e., reduce the standard errors) of the impact estimates for a given sample size (Orr, 1999) and reduce attrition bias from missing data (see Puma, Olsen, Bell, & Price, 2009). The model below produces an estimate of the average treatment impact of ¡Cúidate! across the three replication sites.

Because random assignment occurred at the individual level (not the classroom or school level) within randomization blocks, we estimated the impacts of ¡Cúidate! using a one-level fixed-effects model that included a series of indicator variables representing each of the randomization blocks defined by site, school, semester, and gender (Bloom, 2006, p.13). The model has the basic structure of Equation 1.

$$(1) \quad Y_i = \beta_1 T_i + \sum_{k=1}^K \lambda_k X_{ki} + \sum_{m=1}^M \gamma_{0m} D_m + \varepsilon_i$$

In this model¹¹

¹¹ The analyses presented in this report used linear probability models for binary outcomes. A set of robustness analyses were conducted using multilevel logistic regression models and using multi-level linear models with heteroscedasticity 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.

Y_i is the outcome of interest (e.g., sexual intercourse without birth control) for the i^{th} individual in the m^{th} randomization block;

T_i is an indicator variable equal to 1 if individual i was assigned to the treatment group and 0 otherwise;

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

D_m is the indicator variable representing the m^{th} randomization block; and

ε_i is the usual random error term.

In this model, β_1 represents the average pooled 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, λ_k , 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 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

Equation 1 estimates the impact of access to *¡Cuidate!*. The crucial difference between the treatment group and control group is that they were randomly assigned to receive access to *¡Cuidate!* 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. Most treatment group members received at least one session of the intervention (very few received no sessions); in fact, the majority of treatment group members received at least 75 percent of the program.¹³

3.6.2 Correcting for the 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 the analyses limits the number of hypothesis tests we conduct to draw causal conclusions, thereby mitigating the risk of incorrectly concluding that *¡Cuidate!* was effective.

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, 2008a, p. 70).

¹³ Class rosters were checked after random assignment results were communicated to the sites to verify the group status and determine whether any control group members somehow accessed the intervention (“crossovers”). There were very few crossovers.

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 the 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 domain of which had multiple outcomes (see Exhibit 3.2), we also applied corrections for multiple comparisons. Specifically, we applied a correction described by Benjamini and 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 third, *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), we applied no adjustments to the criterion for statistical significance.¹⁴ For each exploratory test, we applied the traditional criterion of $p < .05$. As noted previously, exploratory analyses should not be used to make causal conclusions about the effectiveness of *¡Cuidate!*. The results from exploratory analyses are reported separately 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 statistically significant results by chance is high. Even if the intervention had no true impact, we would expect five 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 are found, because it is only credible to report an impact in one site—but not in another—if there is a significant difference between the sites. The site-level results in Appendix B are not adjusted for multiple comparisons, and any significant findings reported there should therefore be interpreted with caution.

¹⁴ 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, 2008b) 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.

3.6.4 Subgroup Analyses

In addition to the overall pooled and site-level impacts, we estimated effects for key subgroups of participants—based on age (less than age 15 / age 15 or older), gender (male/female), race/ethnicity (Hispanic, Black, White, Other), and sexual experience at baseline (never sexually active at baseline / ever sexually active at baseline)—and test 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 the model (i.e., Equation 1 above in Section 3.6.1) and testing for the 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 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.

Additional sensitivity analyses and additional analyses conducted to better understand certain site and subgroup findings are described in Appendices D-F.

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 was used in regression models to account for missing covariates. In the dummy-variable adjustment method, missing covariate values are set to a constant, and indicator variables for such values are added to the impact analysis model (Puma et al., 2009).

4. Implementation Findings and Baseline Characteristics

Before presenting the impact results of the study, in this chapter 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.

Implementation of *¡Cuidate!* was guided by fidelity requirements established by OAH at the outset of the grant award. The guidelines allow 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.¹⁵

Key Findings

Across the three replication sites,

- *¡Cuidate!* was implemented as intended, and participants received a majority of the intervention.
- The intended population (majority Hispanic/Latino) was successfully enrolled and served.
- Youth varied in their demographic characteristics, engagement in risk behaviors, and intentions to engage in risky behavior at baseline.

4.1 Program Implementation

As we noted in Section 1.3, a separate report provides a detailed account of the implementation of *¡Cuidate!* in the three replication sites. The 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 *¡Cuidate!* in the future. Here we provide a summary of findings that are directly relevant to the impact findings reported in the next chapters.

¡Cuidate! 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 the core elements and without modifications that threatened those core elements; and attendance was high.

4.1.1 Staff Hiring and Training

Although the grantees gave the staff who administered the program different titles (e.g., prevention specialist versus health educator), they were consistent in the types of experience and skills that they looked for when hiring. Experience working with adolescents and bilingualism, or a Hispanic background, were considered essential. In addition, and given equal emphasis, was comfort in addressing sexual health issues. (One of the grantees required candidates to give a presentation on adolescent sexual health to staff as part of the screening process.) Ultimately, staff hired to deliver the program brought a variety of skills and experience to their roles—their backgrounds included clinical work in reproductive health, tutoring at-risk youth, adolescent mental health, dating violence and sexual assault prevention, and HIV/AIDS outreach work.

¹⁵ A more detailed description and analysis of implementation is provided in the implementation report.

IMPLEMENTATION FINDINGS AND BASELINE CHARACTERISTICS

All of the staff received the official training provided by the curriculum distributor and approved by the developer. Grantees offered additional training and encouraged staff 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 the program model, and provided 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 implementation and continued adherence to the core program model.

For *¡Cuidate!*, the developer provided fidelity monitoring log templates to help collect this information. Health educators were required to complete a fidelity log for each session delivered. In addition, OAH provided observation protocols, to be used by supervisory staff on a regular schedule to assess 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 as 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 grantees implementing the same program model.

Program design slightly varied for each replication site, as we described in Section 3.5.2, but this variation did not affect implementation of the core elements of the program model. Each of the replication sites successfully delivered the intervention to youth with fidelity to the program model. Two grantees, however, discovered that they needed to address the implementation challenge of making the *¡Cuidate!* curriculum relevant to non-Hispanic youth who chose to enroll in the study. In most settings, but especially in school settings, targeting a specific ethnic population is not permitted, for obvious reasons. Although the developer, who had included non-Hispanic youth in her test of the program, believes that the concepts are relevant to youth of any background, the impact of the program on non-Hispanic participants was not reported in the original study. Grantee staff delivering the curriculum had to find ways to translate the cultural references and concepts to make them also relevant to non-Hispanic youth.

4.1.3 Participant Attendance and Engagement

Grantees were required to collect and report youth program attendance (by session) using attendance logs. In all three replications, a majority of youths received at least three-quarters of the sessions offered. The numbers differed slightly by site, with Touchstone having the highest attendance and La Alianza having the lowest.¹⁶ In La Alianza, the average number of sessions attended was 5.6 out of 6, and 73 percent of program participants received at least three-quarters of the sessions offered. For Community Action Partnership, the average number of sessions attended was 7.0 out of 8, with 88 percent of program participants receiving at least three-quarters of the sessions offered. In Touchstone, the average number of sessions attended was 7.5 out of 8, and 96 percent of program participants received at least three-quarters of the sessions offered.

Abt's independent observations and focus group sessions with the youth suggest that, in all replication sites, youth actively participated and acquired new information in the sessions.

¹⁶ In La Alianza, attendance was lower where the program was implemented outside the regular school day or in non-school settings (one after-school setting and two community settings).

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 *¡Cuidate!* overall and for each replication site are presented in Exhibit 4.1 below.¹⁷

Age. At baseline, youth in the study sample were, on average, 14.5 years old; however, the averages varied considerably across the replication sites. In Touchstone, where the program was implemented only in eighth-grade classrooms, the average age of study participants was about 13 years—more than one year younger than the average for the combined sample and between almost two and three years younger than in each of the other two sites.

Gender. Girls constituted more than half of the sample.

Race/Ethnicity. More than 70 percent of study youth were Hispanic, 17 percent were non-Hispanic White, fewer than five percent were non-Hispanic Black, and the remainder classified themselves as non-Hispanic “Other” race. Given *¡Cuidate!*’s focus on Latino culture, it is also important to note the difference in the ethnic mix of the samples across the individual sites. In Community Action Partnership, slightly more than half of the youth were Hispanic, and more than one-third were non-Hispanic White, compared with higher proportions of Hispanic youth and much lower proportions of non-Hispanic White youth in the other two replication sites.

Family Structure and Relationships. Across all three replication sites, more than 90 percent of youth lived with one or both biological parents. Almost half reported feeling very close to and cared for by their fathers; about 60 percent reported they felt close to and cared for by their mothers.

Risk Behaviors. Almost half of the sample had ever used alcohol and a quarter had used marijuana. Fewer than 20 percent had ever smoked cigarettes.

Sexual Activity/Risk Behavior/Consequences. Differences between the sites were significant in the extent to which youth had engaged in sexual activity and sexual risk behavior before they entered the study. Nearly a quarter of the overall sample had ever been sexually active and 17 percent were recently sexually active; however, the Touchstone youths, who were younger than youth in the other sites, were strikingly less sexually experienced: Just seven percent had ever been sexually active, and fewer than four percent were recently sexually active. Youth in La Alianza, who were on average about a year older than the average age of the sample as a whole, consistently reported the highest levels of sexual activity. This same pattern repeated for sexual risk behaviors and consequences. The proportions of youth who had engaged in unprotected sex and had experienced consequences were consistently lowest in the Touchstone sample and highest in the La Alianza sample.

¹⁷ Because of very low attrition, 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 short-term analytic sample and the longer-term analytic sample are shown in Appendix Tables G.1 and G.2, respectively.

IMPLEMENTATION FINDINGS AND BASELINE CHARACTERISTICS

Knowledge/Attitudes/Intentions. Not quite half of the youths (48 percent) in the overall sample demonstrated an accurate understanding of pregnancy risk, whereas a somewhat smaller proportion (39 percent) understood STI risks. Across all three sites, the majority of youth reported supportive attitudes toward using protection. Overall, slightly fewer than one quarter of the sample expressed an intention to engage in oral sex in the next 12 months; almost one-third (30 percent) intended to have sexual intercourse in the same period. The youths in the Touchstone sample (again, younger relative to other sites) were less knowledgeable about pregnancy and STI risks and fewer of them expressed intentions to engage in sex in the next 12 months compared with youth in the other two sites. Across the *¡Cúdate!* sample, the overwhelming majority of youth reported their intention 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	Community Action Partnership	La Alianza	Touchstone	<i>¡Cúdate!</i> Overall	p-Value for the Test of Differences Across Sites ^a
Demographic characteristics					
Age (years)					
Mean	14.99	15.57	13.20	14.45	.000***
Grade	9.99	10.17	8.00	9.23	.000***
Gender					
Female	63.11	58.69	52.12	57.03	.000***
Race/Ethnicity^b					
Hispanic	52.00	81.10	75.48	71.83	.000***
Black	0.89	5.95	5.91	4.72	.000***
White	36.22	10.37	11.55	17.03	.000***
Other	10.89	2.59	7.06	6.42	.000***
Family structure and relationships					
Lives with biological parent/s	92.60	90.65	94.77	92.78	.013*
Feels very close to and cared for by father	42.72	35.90	54.77	45.36	.000***
Feels very close to and cared for by mother	54.05	56.71	65.39	59.57	.000***
Risk behaviors					
Ever smoked cigarettes	25.73	18.73	12.20	17.78	.000***
Ever drank alcohol	62.72	55.73	30.16	47.07	.000***
Ever used marijuana	38.70	30.23	13.54	25.52	.000***
Sexual activity					
Ever sexually active ^c	31.76	40.03	7.11	24.44	.000***
Recently sexually active (in the last 90 days) ^c	22.12	28.75	3.31	16.65	.000***
Sexual intercourse in the last 90 days	16.67	25.24	2.91	13.96	.000***
Oral sex in the last 90 days	16.25	20.50	1.98	11.83	.000***
Anal sex in the last 90 days ^c	2.03	4.74		3.62	.019*
Sexual risk behavior					
Sexual intercourse without birth control in the last 90 days	6.08	6.47	0.66	3.98	.000***
Sexual intercourse without a condom in the last 90 days	8.56	14.83	0.93	7.58	.000***
Oral sex without a condom in the last 90 days	14.90	18.14	1.19	10.36	.000***
Anal sex without a condom in the last 90 days ^c	1.81	2.84		2.42	.276
Consequences of sexual risk behavior					
Ever pregnant or gotten someone pregnant	2.48	3.47	0.13	1.85	.000***

IMPLEMENTATION FINDINGS AND BASELINE CHARACTERISTICS

Measure	Community Action Partnership	La Alianza	Touchstone	<i>¡Cúdate!</i> Overall	p-Value for the Test of Differences Across Sites ^a
Diagnosed with STI in the last 12 months	0.23	1.26	0.00	0.49	.002**
Knowledge, attitudes, and intentions					
Knowledge of pregnancy risk ^d	59.82	53.86	36.42	48.14	.000***
Knowledge of STI risk ^d	52.68	44.75	26.60	39.23	.000***
Attitudes toward protection ^e	3.12	3.13	3.02	3.08	.000***
Intentions to have sexual intercourse ^f	36.05	43.81	15.08	30.05	.000***
Intentions to have oral sex ^f	29.48	33.28	12.30	23.67	.000***
Intentions to use birth control if they were to have sexual intercourse ^f	94.78	90.38	92.31	92.25	.030*
Intentions to use a condom if they were to have sexual intercourse ^f	96.17	93.85	92.49	93.86	.038*

Source: Baseline survey administered prior to randomization.

Notes: Data in this table are based on 1,076–1,885 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 or percentages 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 La Alianza and Community Action Partnership, sexual activity refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in Touchstone.

^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 baseline characteristics of study participants differed significantly across the three replication sites (reflecting the differences in youth populations in those sites), there were few significant differences between those assigned to the treatment group versus those 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 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. There were few significant differences between the two groups (see Appendix Tables G.1 and G.2); variables for which there were differences were subsequently included in the impact models as covariates.

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

In this section of the report, we present findings on behavioral outcomes from both the short-term and longer-term follow-up surveys. The findings presented here reflect our analytic strategy of first conducting confirmatory analyses by examining a key set of outcomes for the pooled sample to produce results that are conclusive about the impacts of *¡Cuidate!*, rather than suggestive. We then conducted additional analyses to explore program effects on other related sexual activity, 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 Chapter 7.

5.1 Confirmatory Analyses of Program Impacts on Key Behavioral Outcomes

The pre-specified confirmatory analyses test the impacts of *¡Cuidate!* on the following key outcomes for the full sample: *recent sexual activity* and *sexual intercourse without birth control* in the short term (six months after baseline); *recent sexual activity* and *sexual intercourse without birth control* in the longer term (18 months after baseline); and *pregnancy* (between the baseline and 18-month follow-up survey). In order to minimize the concern that our confirmatory analyses would miss a behavioral impact that occurred early 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 below).

Confirmatory analyses revealed that *¡Cuidate!* did not have a significant impact on key outcomes in the domains of recent sexual behavior at the short-term follow-up or recent sexual behavior at the longer-term follow-up, but a favorable impact on pregnancy approached statistical significance.

More specifically, there were no impacts on recent sexual activity either at the short-term or at the longer-term follow-up. After 18 months, slightly more than one-quarter of participants in both groups reported engaging in sexual activity within the last 90 days. Similarly, we found no differences on rates of engaging in sexual intercourse without birth control at either time period. Very few participants in either group reported engaging in unprotected sexual intercourse. After six months, 5.8 percent of youth in the treatment group reported having sexual intercourse without birth control in the past 90 days compared with 4.9 percent of youth in the control group. The difference between the two groups was even smaller after 18 months. Neither of these differences between the treatment and control groups was statistically significant.

Key Behavioral Impact Findings

Confirmatory analyses revealed:

- No impact of *¡Cuidate!* on recent sexual activity and sexual intercourse without birth control at the short-term or longer-term
- Favorable impact (approaching statistical significance) on pregnancy at the longer-term

Exploratory analyses revealed no significant overall effects of *¡Cuidate!* on:

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

IMPACTS ON SEXUAL BEHAVIORS AND THEIR CONSEQUENCES

The final confirmatory analysis tested *¡Cuidate!*'s impact on pregnancy in the period between baseline and the longer-term follow-up. After 18 months, there were few pregnancies across the study sample. Only 2.7 percent of youth in the treatment group and 4.4 percent of youth in the control group reported getting pregnant or having gotten someone pregnant since the baseline. While not quite reaching the established criterion for statistical significance of $p < .05$, the finding does favor the treatment group.

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	18.79	17.83	0.96	.516 ^d	27.93	26.59	1.34	.481 ^d
Sexual risk behavior (percentage responding affirmatively)								
Sexual intercourse without birth control (in the last 90 days)	5.77	4.86	0.90	.383 ^d	7.83	7.46	0.36	.776 ^d
Consequences of Sexual Risk Behavior (percentage responding affirmatively)								
Pregnant or gotten someone pregnant since baseline	n/a	n/a	n/a	n/a	2.70	4.38	-1.68	.061 ^e

Source: Follow-up surveys administered six and 18 months after baseline.

Note: Short-term results are based on 2,009–2,012 respondents who provided valid survey responses to relevant items. Longer-term results are based on 1,869–1,870 respondents who provided valid survey responses to relevant items.

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

^b The treatment effect was estimated in a one-level fixed-effects regression 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 La Alianza and Community Action Partnership, sexual activity refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in Touchstone.

^d After application of the 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 also 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 test program effects on other behavioral outcomes related to sexual activity, sexual risk behavior, and consequences for the full sample (see Exhibit 5.2 below).

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

¡Cuidate! had no statistically significant effects on the prevalence of sexual activity at six or 18 months after study enrollment. Slightly fewer than one-quarter of youth in both the treatment and control groups reported that they had sexual intercourse in the last 90 days at the 18-month follow-up; slightly fewer reported having recently engaged in oral sex; and fewer than five percent reported engaging in anal sex.

IMPACTS ON SEXUAL BEHAVIORS AND THEIR CONSEQUENCES

At both the short- and longer-term follow-up, the program similarly had no effect on initiation of sexual activity. By the time of the longer-term follow-up, approximately 28 percent of youth in both groups who were sexually inexperienced at baseline reported having engaged in some sexual activity.

We also found no evidence of program effects on rates of sexual risk behaviors, including sexual intercourse, oral sex, or anal sex without a condom. At the longer-term follow-up, slightly fewer than 15 percent of youth in both groups reported having sexual intercourse without a condom in the past 90 days; approximately 20 percent reported engaging in recent oral sex without a condom, and very few youth (fewer than five percent) in either the treatment or control group reported engaging in anal sex without a condom.

At the time of the longer-term follow-up survey, *¡Cuidate!* had no effect on STI diagnoses. Fewer than one percent of youth in the treatment and control groups reported being diagnosed with an STI in the prior 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	15.48	14.09	1.39	.312	23.52	22.52	1.00	.586
Oral sex in the last 90 days	14.69	13.13	1.56	.266	22.10	20.49	1.62	.368
Anal sex in the last 90 days ^c	2.48	2.87	-0.39	.704	3.70	4.30	-0.59	.646
Initiation of sexual activity ^d	14.62	12.86	1.76	.303	27.84	27.59	0.26	.911
Sexual risk behavior (percentage responding affirmatively)								
Sexual intercourse without a condom (in the last 90 days)	9.81	8.10	1.70	.157	14.38	14.79	-0.41	.799
Oral sex without a condom (in the last 90 days)	12.93	11.25	1.68	.211	20.12	18.86	1.26	.471
Anal sex without a condom (in the last 90 days) ^c	1.46	1.99	-0.53	.525	2.20	3.58	-1.38	.213
Consequences of Sexual Risk Behavior (percentage responding affirmatively)								
Diagnosed with STI in the last 12 months	n/a	n/a	n/a	n/a	0.78	0.95	-0.17	.722

Source: Follow-up surveys administered six and 18 months after baseline.

Note: Short-term results are based on 2,009–2,012 respondents who provided valid survey responses to relevant items, except for the items measuring anal sex ($n=1,173$). Longer-term results are based on 1,869–1,870 respondents who provided valid survey responses to relevant items, except for the items measuring anal sex ($n=1,095$) and pregnancy ($n=1,849$).

IMPACTS ON SEXUAL BEHAVIORS AND THEIR CONSEQUENCES

^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 one-level fixed-effects regression 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 Youth were not asked about anal sex in Touchstone.

^d Sexual activity is defined differently across grantees. In La Alianza and Community Action Partnership, sexual activity refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in Touchstone. The sample size for the initiation of sexual activity outcome at the short term is 1,526, as this outcome includes only youth who were not sexually active at baseline. The sample size at the longer-term follow-up is 1,426.

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

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

¡Cuídate!'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 positive effects would be sustained over time to change youth 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 *¡Cuídate!* affects those non-behavioral outcomes. Specifically, when delivered with fidelity, the program is intended to affect youth 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 *¡Cuídate!* 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 detailed earlier, the analyses in this chapter and the next are exploratory.

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-ups.¹⁸

6.1 Knowledge of Pregnancy and STI Risk

At both the short-term (six months after baseline) and longer-term (18 months after baseline) follow-up, *¡Cuídate!* had statistically significant effects on knowledge of STI risk. Short-term effects on knowledge of pregnancy risk were not sustained at the longer term.

Key Non-Behavioral Findings

At both the short-term and longer-term follow-ups:

- *¡Cuídate!* increased youths' knowledge of STI risk.
- *¡Cuídate!* improved youths' attitudes toward using birth control or condoms.
- *¡Cuídate!* increased students' confidence in their condom negotiation skills.

At 6 months only:

- *¡Cuídate!* increased youths' knowledge of pregnancy risk.

At 18 months only:

- *¡Cuídate!* increased students' confidence in their refusal skills.

The program had no effect on:

- Attitudes toward risky sexual behavior.
- Motivation to delay childbearing.
- Intentions to engage in sexual activity or to use condoms or other birth control.

¹⁸ 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-%C2%A1cu%C3%ADdate>.

EFFECTS ON NONBEHAVIORAL INTERMEDIATE OUTCOMES

In the short term (after six months), treatment group youths knew significantly more about pregnancy risk and STI prevention, transmission, and treatment than did control group youths. After 18 months, there was no significant difference between the two groups in their knowledge of pregnancy risk (i.e., some portion of the control group caught up); however, youth in the treatment group continued to be significantly more knowledgeable about STI risk than were their control group counterparts (Exhibit 6.1).

Exhibit 6.1: Short-Term and Longer-Term Effects of ¡Cuídate! on Knowledge

Outcome	Adjusted Treatment Mean ^a	Unadjusted Control Mean	Treatment Effect ^b	p-Value
Short-Term Follow-Up				
Knowledge of pregnancy risk ^c	67.07	60.95	6.12***	.000
Knowledge of STI risk ^c	63.67	53.01	10.66***	.000
Longer-Term Follow-Up				
Knowledge of pregnancy risk ^c	69.34	68.04	1.30	.323
Knowledge of STI risk ^c	64.84	58.67	6.17***	.000

Source: Follow-up surveys administered six months and 18 months after baseline.

Note: Results in this table are based on 2,022 respondents (short-term survey) and 1,885 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 one-level fixed-effects regression 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

¡Cuídate! improved attitudes toward protection.

Youth in the treatment group reported significantly greater support for the use of birth control and condoms than did youth in the control group at the short-term follow-up, and the effect was sustained over time through the longer-term follow-up period.

At both time points, *¡Cuídate!* had no statistically significant effect on youth attitudes toward risky sexual behavior. Even at baseline, the majority of youths in both the treatment and control groups rejected the view that risky sexual behaviors were acceptable (Exhibit 6.2).

Exhibit 6.2: Short-Term and Longer-Term Effects of ¡Cuídate! 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	3.24	3.14	0.10***	0.24	.000
Attitudes toward risky sexual behavior ^e	3.12	3.33	-0.21	n/a	.692
Longer-Term Follow-Up					
Attitudes toward protection ^d	3.24	3.17	0.07***	0.16	.000
Attitudes toward risky sexual behavior ^e	4.01	4.41	-0.40	n/a	.525

Source: Follow-up surveys administered six months and 18 months after baseline.

EFFECTS ON NONBEHAVIORAL INTERMEDIATE OUTCOMES

Note: Results in this table are based on 2,011–2,022 respondents (short-term survey) and 1,872-1,883 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 one-level fixed-effects regression 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

***¡Cuidate!* had no effect on motivation to delay childbearing.**

At both time points, youth in both the treatment and control groups were highly motivated to delay childbearing. Youth in both groups indicated a strong belief in the importance of delaying childbearing until personal goals have been achieved, and there were no significant differences between the groups (Exhibit 6.3).

Exhibit 6.3: Short-Term and Longer-Term Effects of *¡Cuidate!* 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	3.69	3.69	-0.00	-0.01	.907
Longer-Term Follow-Up					
Motivation to delay childbearing ^d	3.68	3.70	-0.02	-0.03	.563

Source: Follow-up surveys administered six months and 18 months after baseline.

Note: Results in this table are based on 2,015 respondents (short-term survey) and 1,877 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 one-level fixed-effects regression 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

***¡Cuidate!* did not affect youth intentions to engage in sexual activity or to use protection if they were to engage in sexual activity.**

At both time points, youth in both treatment and control groups were equal in their expectations of engaging in sexual intercourse or oral sex in the 12 months after the surveys. Nearly all youth reported their intention to use condoms or birth control if they were to engage in sexual intercourse (Exhibit 6.4).

EFFECTS ON NONBEHAVIORAL INTERMEDIATE OUTCOMES

Exhibit 6.4: Short-Term and Longer-Term Effects of ¡Cuídate! on Intentions

Outcome	Adjusted Treatment Mean ^a	Unadjusted Control Mean	Treatment Effect ^b	p-Value
Short-Term Follow-Up				
Sexual intercourse ^c	40.38	39.07	1.31	.470
Oral sex ^c	37.16	36.60	0.56	.762
Use birth control if they were to have sexual intercourse ^c	93.23	92.42	0.80	.491
Use a condom if they were to have sexual intercourse ^c	92.89	92.74	0.15	.898
Longer-Term Follow-Up				
Sexual intercourse ^c	50.53	51.97	-1.43	.494
Oral sex ^c	46.89	45.57	1.32	.520
Use birth control if they were to have sexual intercourse ^c	90.81	91.56	-0.75	.576
Use a condom if they were to have sexual intercourse ^c	90.78	89.55	1.23	.391

Source: Follow-up surveys administered six months and 18 months after baseline.

Note: Results in this table are based on 1,996–2,005 respondents (short-term survey) and 1,860–1,867 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 one-level fixed-effects regression 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.

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

6.5 Skills

¡Cuídate! had a favorable effect on condom negotiation skills at both time points and an effect on refusal skills after 18 months only.

At the short-term follow-up, ¡Cuídate! had a statistically significant effect on perceived condom negotiation skills. Program participants were more likely to report that they could successfully negotiate condom use with a partner than were control group participants. This effect was sustained at the longer-term follow-up. In addition, at the longer-term follow-up, program participants were more likely than control group youth to perceive they had the necessary skills to refuse unwanted sex (Exhibit 6.5).

Exhibit 6.5: Short-Term and Longer-Term Effects of ¡Cuídate! 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	3.19	3.13	0.06	0.08	.062
Perceived condom negotiation skills ^d	3.53	3.46	0.07***	0.14	.002
Longer-Term Follow-Up					
Perceived refusal skills ^d	3.23	3.15	0.09*	0.11	.012
Perceived condom negotiation skills ^d	3.56	3.49	0.07**	0.14	.004

EFFECTS ON NONBEHAVIORAL INTERMEDIATE OUTCOMES

Source: Follow-up surveys administered six months and 18 months after baseline.

Note: Results in this table are based on 2,015–2,016 respondents (short-term survey) and 1,877–1,878 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 one-level fixed-effects regression 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 in Chapter 5 provide the best evidence to answer with confidence whether or not *¡Cuidate!* 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 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 five percent of the tests conducted to see statistically significant test results even when there were no real effect of the intervention.

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

Key Site and Subgroup Findings

Site-Level Differences:

- At the 18-month follow-up, *¡Cuidate!* had a significant and favorable effect on the incidence of pregnancy in La Alianza, but not in the other two sites.
- There were significant site-level differences in the effects of *¡Cuidate!* on non-behavioral, intermediate outcomes in the short term that did not persist into the longer term.

Subgroup Differences:

- At the short-term follow-up, there were significant differences in the effects of *¡Cuidate!* on sexual behavior, depending on baseline sexual activity and race/ethnicity.
- At the longer-term follow-up, *¡Cuidate!*'s effect on pregnancy was dependent on youth age, baseline sexual activity, and race/ethnicity.
- *¡Cuidate!* was differentially effective on non-behavioral intermediate outcomes for different subgroups of youth at the short-term and longer-term follow-ups.

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 test for site-level differences in effects at both the short-term and longer-term follow-up periods.

7.1.1 Behavioral Outcomes

With one exception, exploratory analyses found no significant site-level differences in program effects on behavioral outcomes.

There were no site-level differences in the effect of *¡Cuidate!* on sexual activity, sexual risk behaviors, and sexual consequences at either the short term or the longer term (see Appendix Exhibits B.1 and B.2)

with one single exception: At the 18-month follow-up, there was a significant difference by site in the effect of *¡Cuidate!* on pregnancy.¹⁹ The program had a significant and favorable effect on the incidence of pregnancy in La Alianza, but not in the other two replication sites. In La Alianza, compared to control group youth, significantly fewer treatment group youth reported having gotten pregnant (or gotten someone pregnant) between baseline and the longer-term (18-month) follow-up.

7.1.2 Non-Behavioral Intermediate Outcomes

At the short-term follow-up, there were significant site-level differences in the effects of *¡Cuidate!* on non-behavioral, intermediate outcomes, but these differences did not persist into the longer term.

After six months, there were significant site-level differences in the effects of *¡Cuidate!* on knowledge, attitudes toward protection, and perceived refusal skills (see Appendix Exhibit B.3). There were positive program effects on knowledge of pregnancy risk in Community Action Partnership and Touchstone (effect was larger in Touchstone), but no effect in La Alianza. There were positive effects on knowledge of STI risk in all sites, but the effects were largest in Touchstone. In addition, there was a positive program effect on attitudes toward protection in Touchstone, but no effect in the other two sites.

Site-level differences in program effects on knowledge and attitudes highlighted program effectiveness in Touchstone; however, the site-level difference in effects on perceived refusal skills suggested that in La Alianza, but not the other two sites, there was a positive program effect on perceived refusal skills at the short-term follow-up.

However, at the longer-term follow-up, none of these short-term differences in program effects by site on non-behavioral intermediate outcomes persisted and no additional site-level differences emerged (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 subgroups of participants. We specifically looked at whether program effects differed by gender, age, race/ethnicity, and 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. Given that *¡Cuidate!* was originally designed for Hispanic youth, we also separately conducted all study analyses using the sample of Hispanic youth only. These findings are presented in Appendix E.

¹⁹ The study's criterion for statistical significance of results of exploratory analyses was $p < .05$. With 18 tests for differences among sites in the effect of *¡Cuidate!* on behavior and consequences outcomes, the expected number of tests that would be significant by chance is $18 \times 0.05 = 0.90$, or almost one test. Therefore, the finding that one test was significant is not inconsistent with a chance finding. Because the finding was on a key outcome, however, we conducted additional analyses to better understand the results. See Appendix D for details.

7.2.1 Behavioral Outcomes

The exploratory analyses examining differences in the effectiveness of *¡Cuidate!* on sexual activity, sexual risk behavior, and consequences found several differences by subgroup.

At the short-term follow-up, there were unintended effects for some subgroups (see Appendix Exhibit C.1), but at the longer-term follow-up, there were favorable impacts for some subgroups (see Appendix Exhibit C.2).

At the short-term follow-up, we found a significant unintended effect of *¡Cuidate!* on sexual intercourse in the last 90 days for youth who were sexually experienced at baseline, but no effect for youth who were sexually inexperienced at baseline. For White youth, but not other racial/ethnic groups, there were statistically significant unintended effects on oral sex and on oral sex without a condom in the last 90 days. In each of these cases, treatment group youth were more likely to engage in the behavior than were control group youth. However, none of these subgroup differences in effects persisted at the longer-term follow-up.

At the longer-term follow-up, however, we did find significant differences by subgroup in the effects of *¡Cuidate!* on pregnancy. For youth who were sexually experienced at baseline, for youth who were 15 years or older at baseline, and for Hispanic youth, the program had a significant favorable effect on pregnancies (i.e., pregnancies were reduced relative to their control group counterparts). Additional information on these effects is presented in Appendix D.

7.2.2 Non-Behavioral Intermediate Outcomes

Exploratory analyses revealed differential effectiveness of *¡Cuidate!* for different subgroups of youth at the short-term and longer-term follow-ups.

At the short-term follow-up, there were scattered subgroup differences in program effects on knowledge, attitudes toward protection, and perceived refusal skills (see Appendix Exhibit C.3). Most of these differences in program effect did not persist over time, but some new differences emerged at the longer-term follow-up (see Appendix Exhibit C.4).

Younger youths in the treatment group demonstrated significantly more accurate knowledge of pregnancy and STI risk and more positive attitudes toward protection than did younger youths in the control group at the six-month follow-up. These effects were not significant for older youth, with the exception of the effect on knowledge of STI risk; in this case, however, the effect was much larger for younger than older youth. These differences in effects by youth age were not observed at the longer-term follow-up.

There was a significant program effect on knowledge of pregnancy risk for females, but not for males, at six months and at 18 months.

At the short-term follow-up, the program had a significant effect on knowledge of STI risk for treatment group youths who were sexually experienced at baseline, and an even stronger effect for those who were not sexually active at baseline. Although this difference in program effect based on sexual experience at baseline was not sustained at the longer-term follow-up, the overall program effect for the combined group of youth, including those who were either sexually inexperienced or experienced at baseline, was sustained at the longer-term follow-up.

At the short-term follow-up, treatment group youth in the “Other” racial/ethnic category, but not those who identified as Hispanic, non-Hispanic Black, or non-Hispanic White, reported significantly stronger perceived ability to refuse unwanted sex compared to their counterparts in the control group. Again, this difference in program effect by race/ethnicity did not persist over time, but the overall program effect for all racial/ethnic groups combined was sustained at the longer-term follow-up.

At the longer-term follow-up, three additional differences in program effects emerged. While there was a positive program effect on attitudes toward protection for the combined group of boys and girls, the subgroup analyses indicated stronger and significant effects for girls, and weaker and non-significant effects for boys. The overall effect on youths’ intentions to use a condom if they were to have sexual intercourse was not statistically significant, but there was a significant positive program effect for youth who were older and those who were sexually experienced at baseline.

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

Our earlier discussion about possible different prevention and intervention program effects (see Section 2.1) raised a question about whether *¡Cuidate!* 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, *¡Cuidate!* served as both a prevention and an intervention program.

- With youth who were sexually inexperienced when the study began, *¡Cuidate!* 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 would be 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 six months and 18 months after baseline) for youth in all three subgroups.

These graphs (see Appendix F, Exhibit F.1) suggest that there is a possible difference in treatment-control differences between youth who were sexually active when the study started versus youth who initiated sex during the study period, potentially reflecting differential prevention and intervention effects.

PROGRAM EFFECTS BY SITE AND SUBGROUP

No consistent differences emerged when looking at descriptive graphs for individual sites. We also note that 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 very strong evaluation of the *¡Cuidate!* program model. The study applied a stringent test to three replications of the same program, implemented by different grantees, in different geographic locations, with somewhat different youth populations, but within the same time frame 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 *¡Cuidate!* had an overall impact on behavior across multiple implementations of the program, although the impact on pregnancy, driven by a significant effect in one site, trended toward significance.

Though *¡Cuidate!* significantly improved and sustained youths' knowledge, attitudes toward protection, and skills, our confirmatory analyses revealed that these positive impacts on non-behavioral, intermediate outcomes did not translate into significant favorable impacts on the five key behavioral outcomes selected to represent the primary goals of this and all other TPP programs. Nor did the program produce overall favorable significant effects on other measures of sexual activity, sexual risk behavior, and adverse consequences (including the behavioral outcomes also measured in the original study). Therefore, we cannot conclude that *¡Cuidate!* was effective.

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

In one of the three replication sites, there was a significant favorable effect in the incidence of pregnancy at the longer-term follow-up.

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 findings that are suggestive and call for additional investigation.

In one of the three replication sites, La Alianza, there was a significant reduction in the incidence of pregnancy among the treatment group, relative to the control group, at the longer-term follow-up. This finding in isolation could be dismissed as a chance finding, given the number of statistical tests conducted. However, the finding is bolstered by subgroup findings for the same outcome: Among Hispanic youth, youth who were sexually experienced at baseline, and older youth, those in the treatment group were significantly less likely to report a pregnancy than were their control group counterparts. Further investigation showed that these subgroup findings are specific to La Alianza. Our further exploration of this effect on pregnancy in La Alianza (see Appendix D), leads us to conclude that this finding is less likely attributable to chance, but suggestive of a favorable program effect on an important outcome in a single site.

What is puzzling is that, for the most part, we were unable to identify behavioral or other likely precursors to explain this effect on pregnancy. We conducted a series of exploratory and mediational analyses (see Appendix D) to look for possible explanations. These analyses did not provide a clear explanation for this

finding. Our battery of measures, though extensive, was not comprehensive. We were unable to identify effects on likely precursors of the pregnancy finding, but we cannot rule out their existence.

Given that these findings suggest some positive results in La Alianza, we are faced with the question: What was different about the La Alianza replication that could explain the program effect in this site? We believe that the explanation does not reside in the implementation of the program. It is true that La Alianza, a newcomer to federal grant programs, struggled during its first pilot year to find partners and to achieve a stable project management structure. Nevertheless, as in the other two replications, qualified and well-trained staff implemented the program with fidelity and were over time generally able to retain participants.

Differences in the populations served may help explain the differential effects of the program in the three sites. Youth in La Alianza were, on average, slightly older and more sexually experienced than youth in the other two sites. The ethnic profile of youth in the La Alianza replication differed from youth in the other two sites. Hispanic youth in Touchstone (AZ) and Community Action Partnership (CA) were primarily of Mexican descent, most of whose families were not recent immigrants. La Alianza served a more diverse Hispanic population that included youth with families from Puerto Rico, the Dominican Republic, Colombia, and Mexico, as well as other Central and South American countries, many of whose families were relative newcomers to the United States.

Taken together with the absence of program impact on the overall (pooled) sample, plus the findings from the original evaluation,²⁰ these findings suggest that the program's positive effects may be limited to Hispanic youth whose families are more recent immigrants. Youth from more settled Mexican-American families may be no more receptive to *¡Cuidate!*'s culturally specific messages than are non-Hispanic youth.

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 program models were intentionally selected to maximize what would 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 *¡Cuidate!* Based on this study, we cannot conclude that *¡Cuidate!* was effective, although exploratory analyses suggest that the program may be effective in some locations with certain subgroups of youth.

²⁰ The sample of the original study, which found favorable effects of *¡Cuidate!* (Villaruel et al., 2006), was primarily Puerto Rican.

References

- American Academy of Pediatrics. (2013). *Policy statement: Condom use by adolescents*. doi: 10.1542/peds.2013-2831.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- 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.
- Bloom, H. S. (2006). *The core analytics of randomized experiments for social research*. MDRC Working Papers on Research Methodology. <http://www.mdrc.org/publications/437/full.pdf>
- Centers for Disease Control and Prevention (CDC). (2004). HIV Surveillance Report, Vol. 15. Atlanta, GA: Author.
- Centers for Disease Control and Prevention (CDC). (2012). *Estimated HIV incidence among adults and adolescents in the US 2007-2010*. HIV Surveillance Report, Vol. 17. Atlanta, GA: Author.
- Constantine, J., Player D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). *An evaluation of teachers trained through different routes to certification, final report* (NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Gleason, P., Clark, M., Tuttle, C., & Dwyer, E. (2010). *The evaluation of charter school impacts: Executive summary* (NCEE 2010-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Goesling, B., Colman, C., Trenholm, C., Terzian, M., & Moore, K. (2014). Programs to reduce teen pregnancy, sexually transmitted infections, and associated risk behaviors: A systematic review. *Journal of Adolescent Health*, 54(5), 499-507.
- Hayes, A. F., Preacher, K. J., & Myers, T. A. (2011). Mediation and the estimation of indirect effects in political communication research. In E. P. Bucy & R. L. Holbert (Eds.), *The sourcebook for political communication research: Methods, measures, and analytical techniques* (pp. 434-465). New York: Routledge. Retrieved from <http://quantpsy.org/pubs.htm>.
- Jemmott III, J. B. (1992). Reductions in HIV risk-associated sexual behaviors among black male adolescents: Effects of an AIDS prevention intervention. *American Journal of Public Health*, 82(3), 372-377.
- Jemmott, J. B., Jemmott, L. S., Fong, G. T., & McCaffree, K. (1999). Reducing HIV risk associated sexual behavior among African American adolescents: Testing the generality of intervention effects. *American Journal of Community Psychology*, 27(2), 161-187.

-
- 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.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.
- Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods*. Thousand Oaks, CA: Sage.
- Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Schochet, P. Z. (2008a). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics, 33*(1), 62-87.
- Schochet, P. Z. (2008b). *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.
- U.S. Census Bureau. (2012). *CB12-FF19*. Washington, DC: Author.
- U.S. Department of Health and Human Services (HHS). (2010). *Teenage pregnancy prevention: Replication of evidence-based programs. Funding opportunity announcement and application instructions*. Washington, DC: Office of Adolescent Health, Office of Public Health and Science.
- Villaruel, A. M., Jemmott, J. B., & Jemmott, L. S. (2006). A randomized controlled trial testing an HIV prevention intervention for Latino youth. *Archives of Pediatrics & Adolescent Medicine, 160*(8), 772-777.
- Villaruel A. M., Jemmott, L. S., & Jemmott III, J. B. (2005). Designing a culturally-based intervention to reduce HIV sexual risk for Latino adolescents. *The Journal of Association of Nurses in AIDS Care, 16*, 23-31.

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, oral sex and anal sex, recent sexual intercourse without birth control, sexual intercourse without a condom, oral sex without a condom and anal 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 those messages in their knowledge, attitudes, motivation, intentions, and skills—all of which are hypothesized precursors of change in youths’ 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 11 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 (described below) 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 La Alianza and Community Action Partnership, <i>sexual activity</i> refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in Touchstone.
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.

APPENDIX A: MEASURES

Measure	Item	Coding
Anal sex in the last 90 days	Now please think about the past 3 months. In the past 3 months, have you had anal sex?	Youth responded to this question with a yes(1)/no(0) answer. Youth who reported engaging in anal sex in the last 90 days received a score of 1 on the measure. Those who reported they had not engaged in anal 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.
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) since baseline and 0 representing no sexual activity since baseline. 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 La Alianza and Community Action Partnership, <i>sexual activity</i> refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in Touchstone.
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? <ul style="list-style-type: none"> • 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.
Anal sex without a condom (in the last 90 days)	In the past 3 months, have you had anal sex without using a condom, even once?	Youth responded to this question with a yes(1)/no(0) answer. Youth who reported engaging in anal 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 anal 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 anal sex in the last 90 days or that they had never been sexually active.

Measure	Item	Coding
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 indicated 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 portend changes in behavior. At the short-term and longer-term follow-ups, 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 youths’ 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
Non-Behavioral Intermediate Outcomes	
Knowledge of pregnancy risk (4 items)	
Used correctly, how much can birth control pills reduce pregnancy risk?	Youth responded to this question on a scale from 1="Not at all" to 4="Completely." This item was recoded into a binary variable where the correct response (a lot) was coded as 1 and all other responses were coded as 0.
Used correctly, how much can condoms reduce pregnancy risk?	Youth responded to this question on a scale from 1="Not at all" to 4="Completely." This item was recoded into a binary variable where the correct response (a lot) was coded as 1 and all other responses were coded as 0.
A couple that has had unprotected sex and not gotten pregnant does not have to worry about getting pregnant.	Youth indicated the veracity of this statement, responding on a scale from 1="I am sure it's true" to 5="I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was false, and 0 indicates they were sure or thought the statement was true or did not know.
A woman is protected from pregnancy the day she begins taking the pill.	Youth indicated the veracity of this statement, responding on a scale from 1="I am sure it's true" to 5="I am sure it's false." This item was recoded into a binary variable where 1 indicates youth were sure or thought the statement was false and 0 indicates they were sure or thought the statement was true or did not know.
Knowledge of STI risk (12 items)	
You can't get infected with HIV if you have sex only once or twice 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.

Component Items	Coding
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 ($\alpha=0.79$).²¹
- ***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 ($\alpha=0.81$).

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.

²¹ 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.”

APPENDIX A: MEASURES

Component Items	Coding
Birth control is important to make sex safer.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward birth control.
Birth control has too many side effects.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." High values indicate more positive attitudes toward birth control.
Using birth control is morally wrong.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." High values indicate more positive attitudes toward birth control.
Condoms are too much trouble to use.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." High values indicate more positive attitudes toward condoms.
Condoms are pretty easy to get.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward condoms.
Condoms are important to make sex safer.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." We reverse coded this item so that higher values indicate more positive attitudes toward condoms.
Using condoms means you don't trust your partner.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." High values indicate more positive attitudes toward condoms.
Using condoms is morally wrong.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." High values indicate more positive attitudes toward condoms.
Condoms decrease sexual pleasure.	Youth expressed their agreement with this statement, responding on a scale from 1="Strongly agree" to 4="Strongly disagree." High values indicate more positive attitudes toward condoms.
Attitudes toward risky sexual 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 youths' 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. This measure is

described below and detailed information about its component items is presented in Exhibit A.4. 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 ($\alpha=0.87$).

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 single items presented in Exhibit A.5 to examine impacts on youths' intended or anticipated sexual behavior 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. We constructed two measures to examine program impacts on youths' 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 ($\alpha=0.87$).
- **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 ¡Cuidate! intervention on all of the outcomes of interest. Although the pooled analysis is the primary focus of this study, study stakeholders clearly were very interested in examining the effects on outcomes 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. The final column of each table shows the statistical result for a test of differences in the treatment effect across sites. Only when a statistically significant difference among sites was found was the corresponding site-specific effect discussed in the main text

Exhibit B.1 Short-Term Effects on Sexual Activity and Sexual Risk Behavior by Site

Outcome	Community Action Partnership (n=486)				La Alianza (n=688)				Touchstone (n=838)				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	26.74	24.26	2.48	.398	28.61	28.97	-0.36	.889	7.12	6.03	1.09	.636	.764
Sexual intercourse in the last 90 days	21.65	17.82	3.83	.161	24.20	24.21	-0.01	.996	5.63	4.60	1.03	.630	.556
Oral sex in the last 90 days	23.24	19.31	3.93	.159	20.90	20.72	0.18	.941	5.25	4.03	1.22	.577	.587
Anal sex in the last 90 days ^d	0.69	1.98	-1.29	.407	3.89	3.59	0.30	.826					.441
Sexual risk behavior (percentage responding affirmatively)													
Sexual intercourse without birth control in the last 90 days	8.81	6.44	2.37	.250	8.68	8.33	0.35	.846	1.89	1.44	0.45	.777	.712
Sexual intercourse without a condom in the last 90 days	13.43	8.91	4.52	.058	16.49	15.08	1.41	.497	2.80	2.59	0.21	.909	.358
Oral sex without a condom in the last 90 days	21.48	16.34	5.14	.053	18.05	17.93	0.12	.957	4.25	3.46	0.79	.703	.313
Anal sex without a condom in the last 90 days ^d	0.39	1.98	-1.59	.210	2.27	1.99	0.28	.803					.267

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

^a This column shows the results for statistical tests of whether the treatment effect varies across 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 (T Effect).

^c The treatment effect was estimated in a one-level fixed-effects regression 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 La Alianza and Community Action Partnership, *sexual activity* refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in Touchstone.

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

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

Outcome	Community Action Partnership (n=448)				La Alianza (n=647)				Touchstone (n=776)				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)													
Recently sexually active (in last 90 days) ^d	39.30	36.36	2.94	.436	36.25	37.50	-1.25	.702	15.35	12.89	2.46	.404	.621
Sexual intercourse in the last 90 days	34.94	32.62	2.32	.524	32.08	32.76	-0.68	.829	10.67	9.12	1.55	.587	.797
Oral sex in the last 90 days	31.67	24.60	7.07*	.048	27.78	29.31	-1.53	.620	12.48	11.64	0.84	.764	.179
Anal sex in the last 90 days ^d	2.84	1.60	1.24	.528	4.49	6.47	-1.98	.247					.216
Sexual risk behavior (percentage responding affirmatively)													
Sexual intercourse without birth control (in last 90 days)	10.10	11.23	-1.13	.657	10.19	9.05	1.14	.605	4.73	4.09	0.64	.746	.783
Sexual intercourse without a condom (in last 90 days)	19.57	22.99	-3.42	.281	21.09	21.98	-0.89	.746	6.54	4.72	1.82	.462	.418
Oral sex without a condom (in last 90 days)	29.81	23.53	6.28	.072	25.57	27.16	-1.59	.599	10.56	10.06	0.50	.853	.219
Anal sex without a condom (in last 90 days) ^d	2.40	1.60	0.80	.637	2.15	5.17	-3.02*	.039					.087
Sexual Consequences (percentage responding affirmatively)													
Pregnant or gotten someone pregnant since baseline	4.68	4.37	0.31	.862	2.66	9.17	-6.51***	.000	1.94	0.94	1.00	.471	.001***
Diagnosed with STI in the last 12 months	-0.17	0.53	-0.70	.463	1.26	2.60	-1.34	.101	1.11	0.00	1.11	.135	.070

Source: Follow-up survey administered 18 months after baseline.

^a This column shows the results for statistical tests of whether the treatment effect varies across 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 (T Effect).

^c 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. 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 La Alianza and Community Action Partnership, *sexual activity* refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in Touchstone.

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

Exhibit B.3: Short-Term Effects on Non-Behavioral Intermediate Outcomes by Site

Outcome	Community Action Partnership (n=490)					La Alianza (n=692)					Touchstone (n=840)					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	77.06	71.43	5.63 *	.019		63.64	62.85	0.79	.707		64.23	53.50	10.73 ***	.000		.002 **
Knowledge of STI risk ^e	69.59	62.11	7.48 ***	.000		64.03	57.08	6.95 ***	.000		60.38	44.79	15.59 ***	.000		.000 ***
Attitudes																
Attitudes toward protection ^f	3.29	3.23	0.06	.064	0.14	3.25	3.21	0.04	.127	0.10	3.20	3.03	0.17 ***	.000	0.42	.001 ***
Attitudes toward risky sexual behavior ^g	4.08	3.80	0.28	.790		2.72	3.40	-0.68	.455		2.87	3.00	-0.13	.877		.780
Motivation																
Motivation to delay childbearing ^f	3.66	3.69	-0.03	.552	-0.05	3.71	3.74	-0.04	.402	-0.07	3.70	3.66	0.04	.299	0.07	.345
Intentions to engage in the following behaviors in the next 12 months																
Intention to have sexual intercourse in the next 12 months ^h	50.71	45.32	5.39	.134		52.58	52.40	0.18	.954		25.36	25.66	-0.30	.917		.419
Intention to have oral sex in the next 12 months ^h	46.57	45.05	1.52	.678		45.81	44.40	1.41	.658		25.23	25.95	-0.72	.802		.843
Intention to use birth control if they were to have sexual intercourse in the next 12 months ^h	94.81	92.54	2.27	.326		91.22	92.37	-1.15	.567		93.88	92.40	1.48	.413		.476
Intention to use a condom if they were to have sexual intercourse in the next 12 months ^h	90.60	93.07	-2.47	.299		92.06	90.00	2.06	.322		94.75	94.52	0.23	.903		.358
Skills																
Perceived refusal skills ^f	3.12	3.20	-0.08	.231	-0.10	3.25	3.10	0.15 **	.007	0.21	3.18	3.11	0.07	.157	0.10	.026 *
Perceived condom negotiation skills ^f	3.51	3.51	0.00	.922	0.01	3.58	3.50	0.09 *	.029	0.17	3.50	3.41	0.10 **	.007	0.19	.248

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

^a This column shows the results for statistical tests of whether the treatment effect varies across 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 one-level fixed-effects regression 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	Community Action Partnership (n=490)					La Alianza (n=692)					Touchstone (n=840)					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	78.61	76.46	2.15	.411		65.77	66.70	-0.93	.682		66.71	64.10	2.61	.203		.476
Knowledge of STI risk ^e	69.88	64.14	5.74**	.003		66.27	62.09	4.18*	.011		61.03	52.96	8.07***	.000		.209
Attitudes																
Attitudes toward protection ^f	3.26	3.24	0.03	.470	0.06	3.31	3.23	0.08*	.011	0.19	3.18	3.09	0.08**	.004	0.19	.426
Attitudes toward risky sexual behavior ^g	7.17	6.38	0.79	.526		2.69	4.29	-1.60	.135		3.19	3.32	-0.13	.895		.325
Motivation																
Motivation to delay childbearing ^f	3.74	3.72	0.02	.711	0.03	3.65	3.69	-0.05	.328	-0.08	3.67	3.69	-0.01	.759	-0.02	.653
Intentions to engage in the following behaviors in the next 12 months																
Intention to have sexual intercourse in the next 12 months ^h	59.34	63.10	-3.76	.367		60.17	59.13	1.04	.774		38.29	40.31	-2.02	.535		.666
Intention to have oral sex in the next 12 months ^h	54.61	53.48	1.13	.782		53.89	51.09	2.80	.430		37.15	36.91	0.24	.941		.865
Intention to use birth control if they were to have sexual intercourse in the next 12 months ^h	91.01	93.55	-2.54	.344		92.17	90.52	1.65	.477		89.53	91.17	-1.64	.437		.429
Intention to use a condom if they were to have sexual intercourse in the next 12 months ^h	89.36	89.84	-0.48	.865		90.69	85.34	5.35*	.030		91.36	92.45	-1.09	.625		.119
Skills																
Perceived refusal skills ^f	3.18	3.17	0.01	.918	0.01	3.28	3.16	0.13*	.032	0.17	3.23	3.13	0.10	.060	0.13	.388
Perceived condom negotiation skills ^f	3.56	3.49	0.08	.113	0.15	3.63	3.53	0.11**	.010	0.21	3.50	3.46	0.03	.375	0.07	.416

Source: Follow-up survey administered 18 months after baseline.

^a This column shows the results for statistical tests of whether the treatment effect varies across 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 (T Effect).

^c The treatment effect was estimated in a one-level fixed-effects regression 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.

Exhibit C.1: Short-Term Effects on Behavioral Outcomes by Subgroup

	Treatment Effect ^a	p-Value ^b
Sexual intercourse in the last 90 days		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,526)	-0.40	.798
Ever sexually active at baseline (n=486)	7.41*	.010
Oral sex in the last 90 days		
Subgroup: Respondent race/ethnicity^c		
Hispanic (n=1,430)	-0.78	.640
Black (n=96)	-1.10	.860
White (n=357)	9.15**	.005
Other (n=126)	7.90	.164
Oral sex without a condom in the last 90 days		
Subgroup: Respondent race/ethnicity^c		
Hispanic (n=1,430)	-0.77	.631
Black (n=96)	-0.18	.975
White (n=357)	8.69**	.006
Other (n=126)	10.12	.061

Source: Follow-up survey administered six 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 ($p < .05$). For example, a test result indicated that the treatment effect on sexual intercourse in the last 90 days was significantly different between youth who were sexually active at baseline and those who weren't.

^a This column shows the estimated treatment effect (treatment-control difference in percentage reporting engaging in the behavior) 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 from zero.

^c 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.

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

APPENDIX C: SUBGROUP EFFECTS

Exhibit C.2: Longer-Term Effects on Behavioral Outcomes by Subgroup

	Treatment Effect ^a	p-Value ^b
Pregnant or gotten someone pregnant since baseline		
Subgroup: Respondent age		
Respondent less than age 15 (<i>n</i> =1,062)	0.23	.848
Respondent age 15 or older (<i>n</i> =787)	-4.26**	.002
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (<i>n</i> =1,421)	-0.59	.559
Ever sexually active at baseline (<i>n</i> =428)	-5.57**	.003
Subgroup: Respondent race/ethnicity ^c		
Hispanic (<i>n</i> =1,332)	-2.84**	.008
Black (<i>n</i> =86)	-3.44	.398
White (<i>n</i> =313)	4.12	.054
Other (<i>n</i> =118)	-3.60	.313

Source: Follow-up survey administered 18 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 ($p < .05$). For example, a test result indicated that the treatment effect on pregnant or gotten someone pregnant was significantly different for younger versus older respondents.

^a This column shows the estimated treatment effect (treatment-control difference in percentage responding affirmatively) 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 from zero.

^c 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.

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

APPENDIX C: SUBGROUP EFFECTS

Exhibit C.3: Short-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=1,144)	8.37***	.000
Respondent age 15 or older (n=878)	3.19	.083
Subgroup: Respondent gender		
Male (n=903)	3.13	.083
Female (n=1,119)	8.56***	.000
Knowledge of STI risk ^c		
Subgroup: Respondent age		
Respondent less than age 15 (n=1,144)	14.17***	.000
Respondent age 15 or older (n=878)	6.05***	.000
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,532)	12.07***	.000
Ever sexually active at baseline (n=490)	5.99**	.001
Attitudes toward protection ^d		
Subgroup: Respondent age		
Respondent less than age 15 (n=1,144)	0.15***	.000
Respondent age 15 or older (n=878)	0.04	.112
Refusal skills ^d		
Subgroup: Respondent race/ethnicity ^e		
Hispanic (n=1,433)	0.07	.055
Black (n=96)	0.23	.107
White (n=359)	-0.14	.065
Other (n=127)	0.37**	.005

Source: Follow-up survey administered six 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 ($p < .05$). 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 from 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.

^e 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.

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

APPENDIX C: SUBGROUP EFFECTS

Exhibit C.4: Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup

	Treatment Effect ^a	p-Value ^b
Knowledge of pregnancy risk ^c		
Subgroup: Respondent gender		
Male (n=810)	-2.64	.188
Female (n=1,075)	4.28*	.014
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,435)	2.72	.069
Ever sexually active at baseline (n=450)	-3.45	.206
Attitudes toward protection ^d		
Subgroup: Respondent gender		
Male (n=809)	0.03	.353
Female (n=1,074)	0.10***	.000
Intention to use a condom if they were to have sexual intercourse ^e		
Subgroup: Respondent age		
Respondent less than age 15 (n=1,065)	-1.94	.304
Respondent age 15 or older (n=800)	5.35*	.014
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,422)	-0.83	.606
Ever sexually active at baseline (n=443)	8.22**	.006

Source: Follow-up survey administered 18 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 ($p < .05$). For example, a test result indicated that the treatment effect on knowledge of pregnancy risk was significantly different between male and female youth.

^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 from 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.

^e Dichotomous variables, reported as percentage of respondents who responded affirmatively.

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

Appendix D: Supplemental Analyses on the Effect on Pregnancy

This appendix describes additional analyses we conducted to better understand the significant effect of *¡Cuidate!* on pregnancy in La Alianza, for Hispanic youth, for older youth, and for youth who were sexually active at baseline. Given that reducing teen pregnancies was the ultimate goal of the program, it was important to more fully understand these site-level and subgroup effects.

First, we aimed to tease apart the site and subgroup effects to determine where the effects truly lie. Given that the cross-section of Hispanic, older, and sexually active youth was most representative of the La Alianza site, we believed the subgroup effects were driven by the La Alianza site. Given the focus of the curriculum on Hispanic youth, we further aimed to distinguish whether the effect was driven by La Alianza or by the Hispanic subgroup, and again concluded that the effect was driven by La Alianza.

Next, after determining that the effect on pregnancy was driven by the La Alianza site, we then conducted several tests to confirm that this finding was not spurious or due to chance, further convincing ourselves of a favorable program effect on pregnancy in La Alianza.

Finally, after confirming the effect on pregnancy in La Alianza, we conducted several mediation analyses in an attempt to explain how the program successfully reduced pregnancies. Our mediation analyses largely suggest that there are unmeasured factors that are largely responsible for the effect on pregnancy.

In the sections that follow, we describe these analyses in more detail.

D.1 Site and Subgroup Effects on Pregnancy

Results presented in the body of the report indicate a favorable program effect on pregnancy in La Alianza, for Hispanic youth, for older youth, and for youth who were sexually active at baseline. Older, Hispanic, and sexually active youth are characteristic of the La Alianza site, and further exploration suggests that the subgroup effects are likely driven by the site-level effect in La Alianza. When we look at the difference between older and younger youth within each of the sites, there are no significant differences in effect on pregnancy (Exhibit D.1 below). However, when we look at the older sample only, we find a significant difference in effect on pregnancy by site. The treatment effect for older youth in La Alianza (estimate=-7.64) was significantly different ($p=.048$; not shown) than the treatment effect for older youth in Community Action Partnership (estimate=-0.89).²²

When we look at the difference between Hispanic and non-Hispanic youth within each of the sites separately, we find no significant differences in effects on pregnancy (see Exhibit D.2 below). However, when we look at the Hispanic-only sample, we find a significant difference in effect on pregnancy by site ($p=.003$) (not shown). More specifically, there is a significant effect only in La Alianza, suggesting La Alianza is driving the effect on pregnancy.

The remaining sections in this appendix describe additional analyses conducted to better understand the favorable program effect on pregnancy in La Alianza.

²² The sample in Touchstone included a very small number of participants who were 15 or older at baseline, so that site is not included in these analyses.

APPENDIX D: SUPPLEMENTAL ANALYSES ON THE EFFECT ON PREGNANCY

Exhibit D.1: Differences in Effect on Pregnancy by Age, within Sites

Site	Youth Less Than Age 15			Youth Age 15 and Older			Test of Differences in Impacts Between Older/Younger ^a
	n	T Effect ^b	p-Value	n	T Effect ^b	p-Value	
La Alianza	117	-3.37	.27	519	-7.64**	.001	ns
Community Action Partnership	211	3.53	.53	229	-0.89	.68	ns

Source: Follow-up survey administered 18 months after baseline.

^a This column shows the results for statistical tests of whether the treatment effect varies between younger and older youth.

^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); ns is not statistically significant.

Exhibit D.2: Hispanic/Non-Hispanic Differences in Effect on Pregnancy, within Sites

Site	Non-Hispanic Youth			Hispanic Youth			Test of Differences in Impacts Between Hispanics and Non-Hispanics ^a
	n	T Effect ^b	p-Value	n	T Effect ^b	P-Value	
La Alianza	117	-1.94	.661	519	-7.00**	.001	ns
Community Action Partnership	211	1.67	.562	229	-2.27	.420	ns
Touchstone	189	1.50	.402	584	0.22	.831	ns

Source: Follow-up survey administered 18 months after baseline.

^a This column shows the results for statistical tests of whether the treatment effect varies between Hispanic and non-Hispanic youth.

^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); ns is not statistically significant.

D.2 Confirming the Effect on Pregnancy in La Alianza

To confirm that the program effect on pregnancy in La Alianza was not a spurious finding explained by other factors, we conducted a number of tests. First, we tested for baseline equivalence between the treatment and control groups within La Alianza (Exhibit D.3 below). The groups were equivalent on all measures, except for grade, the percentage living with a biological parent, and intentions to use a condom if they were to have sexual intercourse.

With the exception of baseline intentions to use a condom if they were to have intercourse, which favored the control group at baseline, each of these variables is included in the impact models, thereby controlling for these treatment-control differences.

Therefore none of these baseline differences accounts for the effect on pregnancy in La Alianza.

APPENDIX D: SUPPLEMENTAL ANALYSES ON THE EFFECT ON PREGNANCY

Exhibit D.3: Characteristics of the Longer-Term Analytic Sample in La Alianza at Baseline

Measure	Treatment Mean ^a	Control Mean	Group Difference ^b	p-Value
Demographic characteristic				
Age				
Mean	15.48	15.49	0.00	.975
Grade				
Mean	10.13	10.00	0.13*	.037
Gender^c				
Female	57.02	57.02	0.00	1.000
Race/ethnicity^d				
Hispanic	82.79	78.30	4.49	.158
Black	5.06	6.38	-1.33	.470
White	9.50	13.19	-3.69	.140
Other	2.65	2.13	0.52	.690
Family structure and relationships				
Lives with biological parents	94.01	86.40	7.60**	.002
Feels very close to and cared for by father	34.67	37.44	-2.78	.507
Feels very close to and cared for by mother	57.49	56.83	0.66	.872
Risk behaviors				
Ever smoked cigarettes	16.89	19.05	-2.16	.475
Ever drank alcohol	56.94	51.95	4.99	.214
Ever used marijuana	30.12	28.70	1.42	.692
Sexual activity				
Ever sexually active ^e	39.04	38.39	0.65	.862
Recently sexually active (in last 90 days) ^e	27.85	26.70	1.15	.744
Sexual intercourse in the last 90 days	23.81	24.89	-1.08	.751
Oral sex in the last 90 days	21.59	17.19	4.40	.179
Anal sex in the last 90 days ^e	4.72	4.52	0.20	.912
Sexual risk behavior				
Sexual intercourse without a condom in the last 90 days	15.29	13.12	2.16	.444
Oral sex without a condom in the last 90 days	19.40	15.38	4.01	.200
Anal sex without a condom in the last 90 days ^e	3.05	2.26	0.79	.575
Sexual intercourse without birth control in the last 90 days	6.97	4.98	1.99	.334
Consequences of sexual risk behavior				
Ever pregnant or gotten someone pregnant (lifetime)	3.45	3.14	0.31	.838
Diagnosed with STI in the last 12 months	1.06	1.33	-0.27	.777
Knowledge^f				
Knowledge of pregnancy risk	53.90	55.82	-1.92	.553
Knowledge of STI risk	44.50	46.12	-1.62	.503
Attitudes^g				
Attitudes toward protection	3.13	3.14	-0.01	.810
Intentions^h				
Intentions to have sexual intercourse	43.30	40.45	2.85	.454
Intentions to have oral sex	34.54	31.19	3.35	.362

APPENDIX D: SUPPLEMENTAL ANALYSES ON THE EFFECT ON PREGNANCY

Measure	Treatment Mean ^a	Control Mean	Group Difference ^b	p-Value
Intentions to use birth control if they were to have sexual intercourse	90.26	91.74	-1.48	.555
Intentions to use a condom if they were to have sexual intercourse	92.62	96.86	-4.24*	.039

Source: Baseline survey administered prior to randomization.

Note: Results in this table are based on the analytic sample of 571–656 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 treatment effects may differ from differences reported between reported means for the treatment and control groups.

^c The analytic model for outcomes estimates impacts within gender groups, and it 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 La Alianza and Community Action Partnership, *sexual activity* refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in Touchstone.

^f Knowledge variables are composite scale scores representing the proportion of items answered correctly.

^g Attitudes variable is a composite scale score, with higher scores indicating 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).

We also looked for any potential gender differences in pregnancies, but found that for both male and female youth, control group members were more likely to get pregnant (or get someone pregnant). Among female youth, 1.8 percent of the treatment group became pregnant, compared to 12.2 percent of the control group. Among male youth, 4.9 percent of the treatment group got someone pregnant, compared to 6.9 percent of the control group. Similarly, we found results to be consistent across randomization blocks.

In addition, although the effect on pregnancy was significant in La Alianza, the number of youth reporting pregnancies between baseline and the longer-term follow-up was small. There were 21 control group members and 15 treatment group members who reported pregnancies. With the finding based on such small numbers of pregnancies, it seemed plausible that very small differences in the number of pregnancies in each group could have led to different conclusions. Therefore, to further validate the pregnancy effect in La Alianza, we explored how sensitive the effect was to small changes in the number of pregnancies in the treatment and control groups.

First, we tested whether the treatment effect would still be significant if one fewer control group member and one additional treatment group member had reported a pregnancy between baseline and follow-up. We found the answer to be yes, the pregnancy effect would still be statistically significant. We also found that the pregnancy effect would still be significant if two fewer control group members and two additional treatment group members reported a pregnancy. If three fewer control group members and three additional treatment group members had reported pregnancies, we found that the treatment effect would not be significant at the $p < .05$ level, although it would still be trending toward significance with $p < .10$.

We interpret the results of this exploration to indicate the pregnancy effect was not very sensitive to small changes in the number of pregnancies in the treatment and control groups; very small changes would not have led to different inferences. Only relatively large differences (e.g., three fewer control group members

APPENDIX D: SUPPLEMENTAL ANALYSES ON THE EFFECT ON PREGNANCY

and three additional treatment group members reporting pregnancies, or more extreme differences) would have produced a result that was not statistically significant. These additional tests further increased our confidence in the favorable effect on pregnancy found in La Alianza.

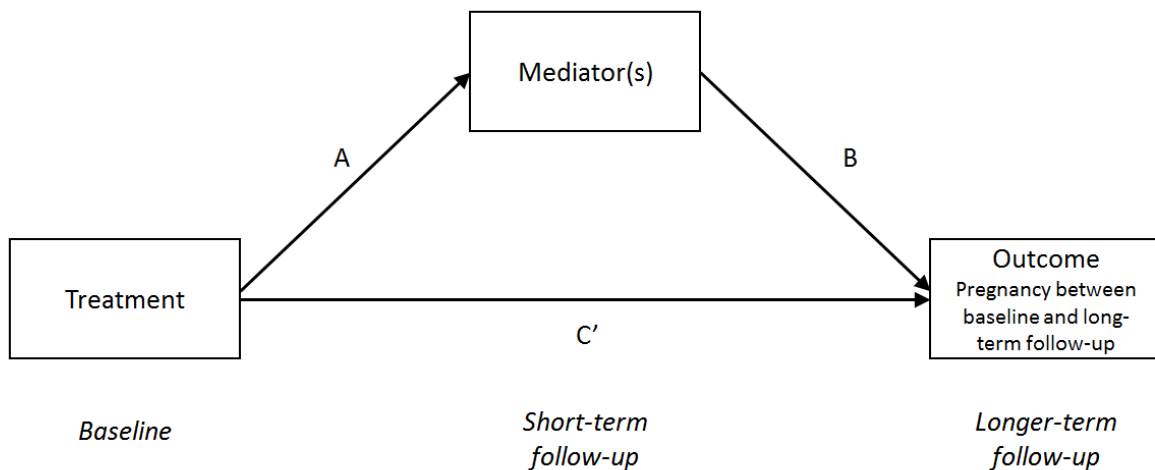
D.3 Explaining the Effect on Pregnancy in La Alianza

As described above, in the La Alianza site, members of the *¡Cúidate!* treatment group were significantly less likely to report having had a pregnancy (or having caused a pregnancy) between baseline and the longer-term follow-up than their control group counterparts. Observing this favorable effect leads inevitably to the question, *By what mechanism did ¡Cúidate! reduce pregnancies in the La Alianza site?*

To help answer this question, we conducted a mediation analysis to test whether treatment-induced changes in pregnancy-related knowledge, attitudes, intentions, skills, or behavior at the short-term follow-up account for (i.e., “mediate”) the observed experimental impact on pregnancy at the longer-term follow-up. In this section, we describe these analyses and present results.

Exhibit D.4 presents a conceptual schematic of mediation. In mediation, an initial variable gives rise to an outcome by way of an intervening variable, or “mediator.” For the current study, we posit that the treatment (the *¡Cúidate!* intervention) had an effect on pregnancy at the longer-term follow-up (the outcome) *through* its effect on knowledge, attitudes, intentions, skills, or behavior at the short-term follow-up (the proposed mediators).

Exhibit D.4: A Simple Mediation Model



Traditional guidance suggests that for mediation to take place, four conditions must generally be met (Baron & Kenny, 1986). This is known as the “causal steps” method.

Step 1: The pathway from the initial variable (random assignment to the treatment) to the outcome must be statistically significant (Path C, coefficient c).

Step 2: The pathway from the initial variable to the proposed mediator must be statistically significant (Path A, coefficient a).

APPENDIX D: SUPPLEMENTAL ANALYSES ON THE EFFECT ON PREGNANCY

Step 3: The pathway from the proposed mediator to the outcome must be statistically significant (Path B, coefficient b).

Step 4: Finally, when the proposed mediator is included as a covariate in a model estimating the relation between the initial variable and the outcome (Path C', coefficient c'), the size and significance of the coefficient is reduced relative to the initial model (coefficient c' < coefficient c).

More recent guidance suggests that some of these conditions can be relaxed to enable detection of a variety of indirect effects including mediation, suppression, confounding, and inconsistent mediation (McKinnon, 2008; Hayes, Preacher, & Myers, 2011).²³ Advances in mediation modeling provide an alternative means of calculating the indirect effect using the product of the Path A and Path B coefficients, known as the “product of coefficients” method. This technique also enables the examination and comparison of several mediators simultaneously in a multiple mediator model. In the present analysis, we used Mplus software (version 7.4) to apply the more recent guidance to the mediation analysis. La Alianza participants with data at baseline and the short-term follow-up or the longer-term follow-up (or both) were included in the analysis ($n=712$). Missing data were addressed using full information maximum likelihood estimation.

We analyzed two sets of potential mediators of pregnancy outcome at La Alianza. The first set comprised a set of potential mediators that were hypothesized by the study team to be the most likely routes or potential mechanisms through which *¡Cuidate!* could have had an effect on pregnancy at the long-term follow-up. This set included the following potential mediators measured at the short-term follow-up:

- Sexual intercourse without birth control;²⁴
- Use of less effective birth control during last intercourse;
- Lifetime sexual intercourse with five or more partners;
- Knowledge concerning leaving space at the end of a condom; and
- Knowledge concerning how to unroll a condom correctly.

The results of the mediation analysis do not support the hypothesis that the potential mediator variables in the set above were mechanisms for the observed treatment effect on pregnancy. For none of the variables in that set was there a significant treatment effect on the potential mediator (Path A). Nor were there significant indirect effects through any of the potential mediators individually (i.e., no specific indirect effects) or through their cumulative or combined effect (i.e., no total indirect effect) as estimated using the “product of coefficients” method in a multiple mediator model.

For the second set of potential mediators, we examined the short-term impact results from La Alianza to identify any measures of program receipt, knowledge, attitudes, skills, or behaviors for which significant treatment effects were found (Path A). We then conducted preliminary analyses to identify the variables within that set that had significant associations with the pregnancy outcome (Path B). The resulting set is shown in the bulleted list below:

²³ Note, mediation is one type of indirect effect among many.

²⁴ We also examined a measure of the frequency of sexual intercourse without birth control.

APPENDIX D: SUPPLEMENTAL ANALYSES ON THE EFFECT ON PREGNANCY

- Receipt of program information:
 - Learned about abstinence from sex;
 - Learned about birth control methods;
 - Learned where to obtain birth control; and
 - Learned how to say no to sex.
- Attitudes toward protection:
 - Attitude that condoms do not decrease sexual pleasure.

We then entered this set into a multiple mediation analysis. The results of this analysis indicate a marginally significant ($p < .10$) total indirect multiple mediation effect of the variables in this set. The results indicate that of the total treatment effect on pregnancy of about 6.5 percentage points, a little more than one percentage point of the total effect was obtained via the total indirect effect of this set of variables (Exhibit D.5). Nor were there any statistically significant effects of the individual proposed mediator variables on the outcome (i.e., no significant specific indirect effects). Because the multiple mediation effect was only marginally significant and explained only a small portion of the total treatment effect on pregnancy, we interpret this result as weak evidence of a partial mediation effect through the variables in this set. These results leave much of the treatment effect on pregnancy unaccounted for.

Because *¡Cuidate!* was originally designed specifically for Hispanic youth, the research team examined whether the indirect effects were present for Hispanic youth or non-Hispanic youth. To do so, we ran a multiple group analysis estimating indirect effects for Hispanics and non-Hispanic subgroups and found no significant indirect effects for either group.

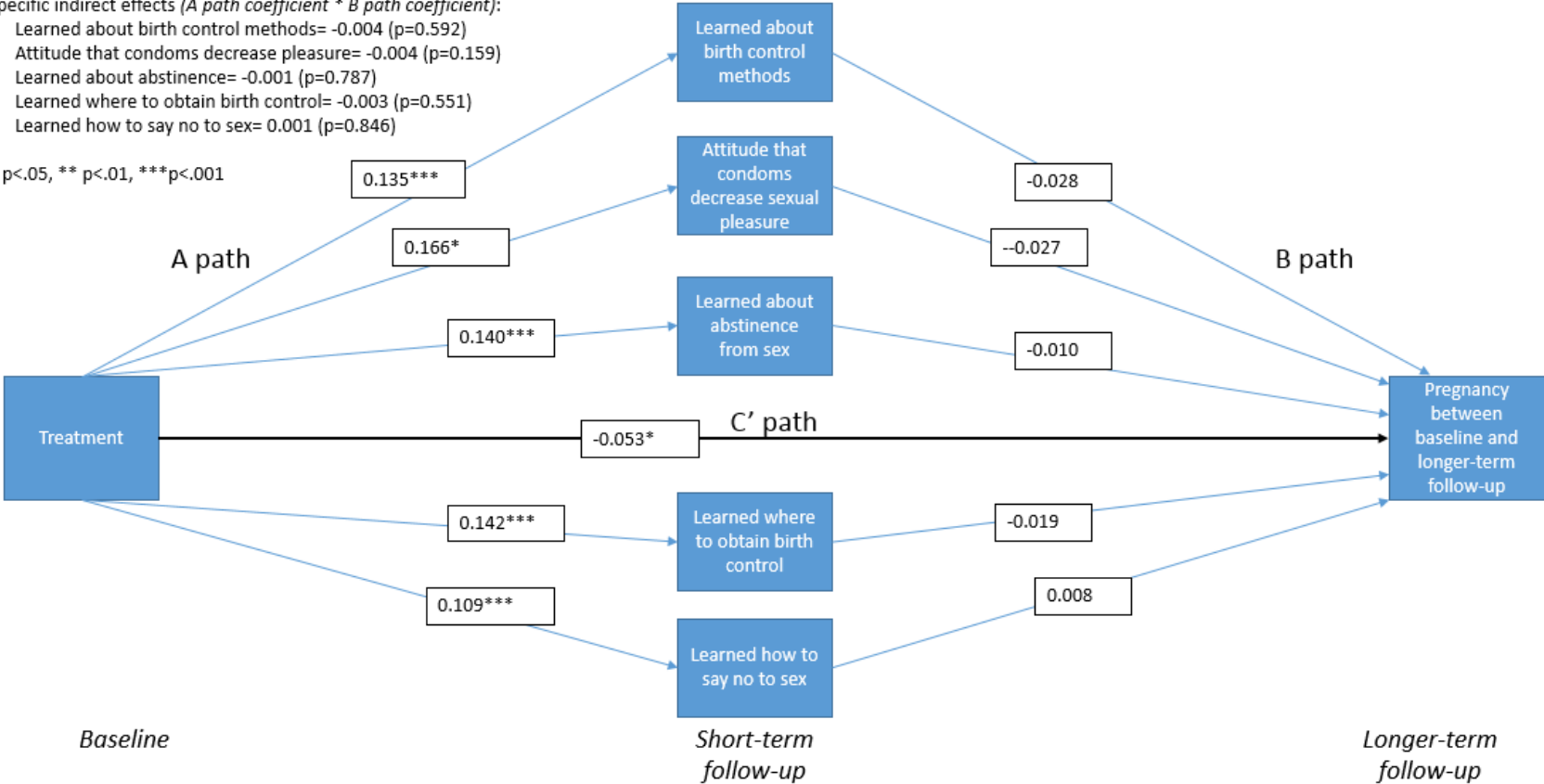
We conclude from these analyses that we lack evidence to support the hypothesis that the impact on pregnancy at La Alianza was mediated via intervention impacts on the pregnancy-related knowledge, attitudes, intentions, skills, or behaviors that were measured at the short-term follow-up. We must conclude that the effect on pregnancy was mediated via other, unmeasured factors.

APPENDIX D: SUPPLEMENTAL ANALYSES ON THE EFFECT ON PREGNANCY

Exhibit D.5: Summary of Results of Multiple Mediation Model

Total effect: -0.065** (p=0.002);
 Direct effect (C' path coefficient): -0.053* (p=0.011)
 Total indirect effect: -0.012 (p=0.053)
 Specific indirect effects (A path coefficient * B path coefficient):
 Learned about birth control methods= -0.004 (p=0.592)
 Attitude that condoms decrease pleasure= -0.004 (p=0.159)
 Learned about abstinence= -0.001 (p=0.787)
 Learned where to obtain birth control= -0.003 (p=0.551)
 Learned how to say no to sex= 0.001 (p=0.846)

* p<.05, ** p<.01, ***p<.001



Appendix E: Sensitivity Analyses Conducted on Hispanic-Only Sample

¡Cuidate! is specifically designed for Hispanic youth, broadly defined, although the developer maintains that its messages are appropriate for youth of all ethnicities because it is solidly based on another evidence-based program (*Be Proud! Be Responsible!*) that is intended for a broader population. While subgroup analyses described in the main report revealed differences between racial/ethnic groups only on refusal skills, oral sex, and oral sex without a condom at the short-term follow-up (no significant effect for Hispanic youth) and pregnancy at the longer-term follow-up (significant favorable effect for Hispanic youth, which we are attributing to a site-level effect at La Alianza), we conducted additional sensitivity analyses by running the full set of analyses on Hispanic youth only (see Exhibits E.1 through E.11).

The results for Hispanic youth only in Exhibits E.1 through E.3 below, largely parallel the findings for the overall pooled sample (shown in the main report). The few exceptions are for treatment effects that were borderline significant for the full sample but significant for the Hispanic-only sample (i.e., perceived refusal skills at the short-term and pregnancy at the longer-term follow-up), or that were significant for the full sample but borderline significant for the Hispanic-only sample (i.e., perceived condom negotiation skills at the longer-term follow-up).

Some differences also emerge when comparing the site-level and subgroup findings for the full sample (shown in the main report) and the Hispanic-only sample (see Exhibits E.4 through E.11).

Exhibit E.1: Short-Term and Longer-Term Effects on Sexual Activity, Sexual Risk Behavior, and Consequences for Hispanic Sample Only

Outcome	Short-Term Impacts				Longer-Term Impacts			
	Adj. T Mean ^a	Unadj. C Mean	T Effect ^b	p-Value	Adj. T Mean ^a	Unadj. C Mean	T Effect ^b	p-Value
Sexual Behavior								
Sexual activity (percentage responding affirmatively)								
Recently sexually active (in the last 90 days) ^c	17.44	17.93	-0.50	.777	25.66	26.41	-0.75	.739
Sexual intercourse in the last 90 days	14.71	14.92	-0.21	.900	21.91	21.94	-0.03	.989
Oral sex in the last 90 days	12.28	13.55	-1.27	.449	19.81	20.39	-0.58	.787
Anal sex in the last 90 days ^c	2.46	2.97	-0.51	.698	3.35	5.05	-1.71	.290
Initiation of sexual activity ^d	12.48	12.00	0.48	.811	25.15	26.52	-1.37	.612
Sexual risk behavior (percentage responding affirmatively)								
Sexual intercourse without birth control (in the last 90 days)	5.73	5.33	0.40	.754	7.86	7.57	0.28	.854
Sexual intercourse without a condom (in the last 90 days)	9.43	8.88	0.55	.711	13.80	14.95	-1.15	.548
Oral sex without a condom (in the last 90 days)	10.52	11.59	-1.07	.502	17.39	18.45	-1.06	.610
Anal sex without a condom (in the last 90 days) ^c	1.25	1.98	-0.73	.487	1.77	3.97	-2.20	.096
Sexual Consequences (percentage responding affirmatively)								
Pregnant or gotten someone pregnant since baseline					1.97	5.08	-3.12**	.004
Diagnosed with STI in the last 12 months					0.69	1.17	-0.47	.429

Source: Follow-up surveys administered six and 18 months after baseline.

Note: Short-term results are based on 1,430–1,433 respondents who provided valid survey responses to relevant items, except for the items measuring anal sex ($n=804$). Longer-term results are based on 1,344–1,346 respondents who provided valid survey responses to relevant items except for the items measuring anal sex ($n=761$) and pregnancy ($n=1,332$).

^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).

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

^b The treatment effect was estimated in a one-level fixed-effects regression 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 reported between reported means for the treatment and control groups.

^c Sexual activity is defined differently across grantees. In La Alianza and Community Action Partnership, sexual activity refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in Touchstone.

^d The sample size for the initiation of sexual activity outcome at the short-term follow-up is 1,104, as this outcome includes only youth who were not sexually active at baseline. The sample size at the longer-term is 1,036.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests). (For the confirmatory outcomes, statistical significance at $p < .05$, $p < .01$, and $p < .001$ implies statistical significance at these levels after applying a Benjamini-Hochberg adjustment for multiple comparisons within the same domain.)

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.2: Short-Term Effects on Non-Behavioral Intermediate Outcomes for Hispanic Sample Only

Outcome	Adjusted Treatment Mean ^a	Unadjusted Control Mean	Treatment Effect ^b	SES ^c	p-Value
Knowledge^d					
Knowledge of pregnancy risk	64.37	57.38	6.98***		.000
Knowledge of STI risk	62.12	51.23	10.88***		.000
Attitudes					
Attitudes toward protection ^e	3.21	3.11	0.10***	0.24	.000
Attitudes toward risky behavior ^f	2.59	3.05	-0.46		.420
Motivation^e					
Motivation to delay childbearing	3.69	3.69	0.00	0.01	.900
Intentions (to engage in the following behaviors in the next 12 months)^g					
Sexual intercourse	38.23	38.31	-0.08		.970
Oral sex	33.97	33.57	0.40		.858
Use birth control if they were to have sexual intercourse	93.52	92.65	0.87		.530
Use a condom if they were to have sexual intercourse	93.16	92.32	0.84		.554
Skills^e					
Perceived refusal skills	3.20	3.12	0.08*	0.11	.036
Perceived condom negotiation skills	3.52	3.45	0.07**	0.15	.007

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

Note: Results in this table are based on 1,423–1,440 respondents 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 one-level fixed-effects regression 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 dichotomous or measured on a 0 to 100 scale, the SES is the treatment effect divided by the pooled standard deviation of the treatment and control groups.

^d Scores represent the average percentage of items answered correctly.

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

^f Score represents the average percentage of items agreed with.

^g Dichotomous variables, reported as percentage of respondents who responded affirmatively.

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

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.3: Longer-Term Effects on Non-Behavioral Intermediate Outcomes for Hispanic Sample Only

Outcome	Adjusted Treatment Mean ^a	Unadjusted Control Mean	Treatment Effect ^b	SES ^c	p-Value
Knowledge^d					
Knowledge of pregnancy risk	67.54	65.44	2.10		.194
Knowledge of STI risk	63.47	57.24	6.23***		.000
Attitudes					
Attitudes toward protection ^e	3.24	3.15	0.09***	0.20	.000
Attitudes toward risky behavior ^f	3.06	3.97	-0.91		.191
Motivation^e					
Motivation to delay childbearing	3.70	3.70	-0.00	-0.01	.880
Intentions (to engage in the following behaviors in the next 12 months)^g					
Sexual intercourse	47.63	50.39	-2.76		.274
Oral sex	42.12	42.16	-0.04		.988
Use a condom if they were to have sexual intercourse	91.21	90.87	0.34		.836
Use birth control if they were to have sexual intercourse	90.27	91.42	-1.15		.484
Skills^e					
Perceived refusal skills	3.24	3.16	0.08*	0.11	.046
Perceived condom negotiation skills	3.55	3.50	0.05	0.11	.065

Source: Follow-up survey administered 18 months after baseline.

Note: Results in this table are based on 1,337–1,354 respondents 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 one-level fixed-effects regression 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 dichotomous or measured on a 0 to 100 scale, the SES is the treatment effect divided by the pooled standard deviation of the treatment and control groups.

^d Scores represent the average percentage of items answered correctly.

^e This construct averages responses ranging from 1 to 4. Higher scores indicate higher levels of the outcome.

^f Score represents the average percentage of items agreed with.

^g Dichotomous variables, reported as percentage of respondents who responded affirmatively.

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

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.4: Short-Term Effects on Sexual Activity and Sexual Risk Behavior by Site for Hispanic Sample Only

Outcome	Community Action Partnership (n=253)				La Alianza (n=552)				Touchstone (n=628)				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 the last 90 days ^d	21.41	23.81	-2.40	.560	29.55	29.15	0.40	.887	5.99	6.95	-0.96	.588	1.89
Sexual intercourse in the last 90 days	17.93	19.05	-1.12	.773	24.93	25.63	-0.70	.795	4.88	5.02	-0.14	.303	2.47
Oral sex in the last 90 days	17.76	21.90	-4.14	.288	19.65	20.20	-0.55	.839	4.08	5.04	-0.96	.853	3.48
Anal sex in the last 90 days ^d	-0.21	2.86	-3.07	.184	3.66	3.03	0.63	.697					.189
Sexual risk behavior (percentage responding affirmatively)													
Sexual intercourse without birth control in the last 90 days	8.01	8.57	-0.56	.849	9.16	8.04	1.12	.588	1.89	1.93	-0.04	.984	.872
Sexual intercourse without a condom in the last 90 days	9.84	11.43	-1.59	.646	17.56	15.08	2.48	.303	2.47	3.09	-0.62	.783	.527
Oral sex without a condom in the last 90 days	15.52	19.05	-3.53	.340	16.19	16.67	-0.48	.853	3.48	4.65	-1.17	.628	.792
Anal sex without a condom in the last 90 days ^d	-0.76	2.86	-3.62*	.048	1.99	1.52	0.47	.714					.067

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

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

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

^c The treatment effect was estimated in a one-level fixed-effects regression 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 reported between reported means for the treatment and control groups.

^d Sexual activity is defined differently across grantees. In La Alianza and Community Action Partnership, sexual activity refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in Touchstone.

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

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.5: Longer-Term Effects on Sexual Behavior, Sexual Risk Behavior, and Consequences by Site for Hispanic Sample Only

Outcome	Community Action Partnership (n=234)				La Alianza (n=527)				Touchstone (n=585)				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)													
Recently sexually active in the last 90 days ^d	34.18	35.11	-0.93	.862	35.58	39.89	-4.31	.233	15.16	12.61	2.55	.459	.391
Sexual intercourse in the last 90 days	31.48	30.85	0.63	.902	31.08	34.97	-3.89	.263	11.59	8.40	3.19	.335	.335
Oral sex in the last 90 days	27.84	23.40	4.44	.378	26.73	30.05	-3.32	.333	11.57	11.76	-0.19	.953	.440
Anal sex in the last 90 days ^d	2.49	2.13	0.36	.901	3.89	6.56	-2.67	.171					.380
Sexual risk (percentage responding affirmatively)													
Sexual intercourse without birth control in the last 90 days	12.12	11.70	0.42	.908	9.69	9.84	-0.15	.953	4.82	4.20	0.62	.795	.975
Sexual intercourse without a condom in the last 90 days	17.16	22.34	-5.18	.253	21.24	24.04	-2.80	.363	7.04	5.04	2.00	.494	.328
Oral sex without a condom in the last 90 days	24.33	21.28	3.05	.534	24.47	27.87	-3.40	.309	9.43	10.08	-0.65	.837	.546
Anal sex without a condom in the last 90 days ^d	1.75	2.13	-0.38	.870	1.87	4.92	-3.05	.057					.345
Sexual Consequences (percentage responding affirmatively)													
Pregnant or gotten someone pregnant since baseline	4.44	6.52	-2.08	.419	2.85	10.44	-7.59***	.000	0.90	0.42	0.48	.770	.003**
Diagnosed with STI in the last 12 months	-0.06	0.00	-0.06	.965	1.41	3.30	-1.89*	.049	0.64	0.00	0.64	.486	.156

Source: Follow-up survey administered 18 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 regression 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 La Alianza and Community Action Partnership, sexual activity refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in Touchstone.

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

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.6: Short-Term Effects on Non-Behavioral Intermediate Outcomes by Site for Hispanic Sample Only

Outcome	Community Action Partnership (n=255)				La Alianza (n=555)				Touchstone (n=630)				p-Value for the Test of Differences Across Sites ^a
	Adj. T Mean ^b	Unadj. C Mean	T Effect ^c	p-Value	Adj. T Mean ^b	Unadj. C Mean	T Effect ^c	p-Value	Adj. T Mean ^b	Unadj. C Mean	T Effect ^c	p-Value	
Knowledge^d													
Knowledge of pregnancy risk	70.41	63.33	7.08*	.040	62.98	62.44	0.54	.822	63.28	51.15	12.13***	.000	.002**
Knowledge of STI risk	63.62	56.90	6.72**	.006	64.57	57.45	7.12***	.000	59.89	44.22	15.67***	.000	.000***
Attitudes													
Attitudes toward protection ^e	3.25	3.13	0.11**	.010	3.26	3.22	0.03	.280	3.16	3.01	0.15***	.000	.022*
Attitudes toward risky behavior ^f	1.50	2.99	-1.49	.262	2.49	3.10	-0.61	.507	3.15	3.03	0.12	.889	.584
Motivation^e													
Motivation to delay childbearing	3.65	3.64	0.01	.883	3.70	3.77	-0.07	.145	3.70	3.64	0.05	.263	.183
Intentions (to engage in the following behaviors in the next 12 months)^g													
Sexual intercourse	44.51	43.81	0.70	.891	52.22	51.02	1.20	.737	24.25	26.27	-2.02	.549	.789
Oral sex	35.10	40.00	-4.90	.342	44.27	39.80	4.47	.214	25.64	26.17	-0.53	.875	.300
Use a condom if they were to have sexual intercourse	91.68	89.52	2.16	.515	91.29	90.36	0.93	.686	94.78	94.96	-0.18	.934	.832
Use birth control if they were to have sexual intercourse	94.31	89.42	4.89	.126	92.20	94.42	-2.22	.318	93.93	92.61	1.32	.529	.171
Skills^e													
Perceived refusal skills	3.04	3.13	-0.09	.318	3.31	3.12	0.19**	.003	3.16	3.11	0.05	.404	.035*
Perceived condom negotiation skills	3.43	3.43	0.00	.980	3.60	3.51	0.09*	.047	3.48	3.40	0.08*	.046	.494

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

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

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

^c The treatment effect was estimated in a one-level fixed-effects regression 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 reported between reported means for the treatment and control groups.

^d Scores represent the average percentage of items answered correctly.

^e This construct averages responses ranging from 1 to 4. Higher scores indicate higher levels of the outcome.

^f Score represents the average percentage of items agreed with.

^g Dichotomous variables, reported as percentage of respondents who responded affirmatively.

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

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.7: Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Site for Hispanic Sample Only

Outcome	Community Action Partnership (n=234)				La Alianza (n=532)				Touchstone (n=588)				p-Value for the Test of Differences Across Sites ^a
	Adj. T Mean ^b	Unadj. C Mean	T Effect ^c	p-Value	Adj. T Mean ^b	Unadj. C Mean	T Effect ^c	p-Value	Adj. T Mean ^b	Unadj. C Mean	T Effect ^c	p-Value	
Knowledge^d													
Knowledge of pregnancy risk	75.00	68.88	6.12	.110	65.36	66.98	-1.62	.531	66.74	62.92	3.82	.121	.161
Knowledge of STI risk	66.84	59.75	7.09**	.010	66.45	62.77	3.68*	.047	60.21	52.01	8.20***	.000	.198
Attitudes													
Attitudes toward protection ^e	3.23	3.18	0.06	.278	3.34	3.24	0.10**	.004	3.16	3.07	0.08*	.011	.763
Attitudes toward risky behavior ^f	4.26	3.95	0.31	.851	2.53	4.66	-2.13	.056	3.14	3.44	-0.30	.778	.357
Motivation^e													
Motivation to delay childbearing	3.70	3.73	-0.03	.739	3.67	3.72	-0.05	.295	3.72	3.68	0.05	.325	.342
Intentions (to engage in the following behaviors in the next 12 months)^g (%)													
Sexual intercourse	52.07	57.45	-5.38	.368	59.98	60.22	-0.24	.952	36.22	40.17	-3.95	.305	.714
Oral sex	44.13	46.81	-2.68	.649	51.71	48.89	2.82	.482	33.64	35.17	-1.53	.688	.650
Use a condom if they were to have sexual intercourse	88.34	91.49	-3.15	.415	92.44	86.41	6.03*	.021	90.68	94.09	-3.41	.173	.021*
Use birth control if they were to have sexual intercourse	87.93	90.32	-2.39	.541	93.61	91.85	1.76	.504	88.21	91.53	-3.32	.189	.358
Skills^e													
Perceived refusal skills	3.20	3.17	0.02	.797	3.31	3.19	0.12	.060	3.19	3.13	0.07	.283	.670
Perceived condom negotiation skills	3.55	3.47	0.08	.230	3.64	3.57	0.07	.106	3.48	3.46	0.02	.619	.640

Source: Follow-up survey administered 18 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 one-level fixed-effects regression 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 reported between reported means for the treatment and control groups.

^d Scores represent the average percentage of items answered correctly.

^e This construct averages responses ranging from 1 to 4. Higher scores indicate higher levels of the outcome.

^f Score represents the average percentage of items agreed with.

^g Dichotomous variables, reported as percentage of respondents who responded affirmatively.

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

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.8: Short-Term Effects on Sexual Behavior and Sexual Risk Behavior by Subgroup for Hispanic Sample Only

Outcome	Treatment Effect ^a	p-Value ^b
Oral sex in the last 90 days		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,101)	0.47	.805
Ever sexually active at baseline (n=329)	-8.14*	.025
Oral sex without a condom in the last 90 days		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,101)	0.99	.581
Ever sexually active at baseline (n=329)	-9.79**	.004

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

Note: Impact estimates for subgroups are shown only if a test for differences in impacts among the subgroups met the study criterion for statistical significance ($p < .05$). For example, a test result indicated that the treatment effect on oral sex was significantly different between groups based on sexual experience at baseline.

^a This column shows the estimated treatment effect (treatment/control difference in percentage reporting engaging in the behavior) 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 from zero.

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

Exhibit E.9: Longer-Term Effects on Sexual Behavior, Sexual Risk Behavior, and Sexual Consequences by Subgroup for Hispanic Sample Only

Outcome	Treatment Effect ^a	p-Value ^b
Sexual intercourse without a condom		
Subgroup: Gender		
Male (n=567)	-5.71	.057
Female (n=779)	2.00	.422
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,036)	2.21	.306
Ever sexually active at baseline (n=310)	-13.30**	.001
Pregnant or gotten someone pregnant since baseline		
Subgroup: Respondent age		
Respondent less than age 15 (n=814)	-0.33	.814
Respondent age 15 or older (n=518)	-7.38***	.000
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,033)	-1.37	.259
Ever sexually active at baseline (n=299)	-9.69***	.000

Source: Follow-up survey administered 18 months after baseline.

Note: Impact estimates for subgroups are shown only if a test for differences in impacts among the subgroups met the study criterion for statistical significance ($p < .05$). For example, a test result indicated that the treatment effect on sexual intercourse without a condom was significantly different for male and female youth.

^a This column shows the estimated treatment effect (treatment/control difference in percentage responding affirmatively) 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 from zero.

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

APPENDIX E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.10: Short-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup for Hispanic Sample Only

Outcome	Treatment Effect ^a	p-Value ^b
Knowledge of pregnancy risk^c		
Subgroup: Respondent gender		
Male (n=630)	2.84	.209
Female (n=810)	9.74***	.000
Subgroup: Respondent age		
Respondent less than age 15 (n=867)	9.35***	.000
Respondent age 15 or older (n=573)	2.86	.223
Knowledge of STI risk^c		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,109)	12.09***	.000
Ever sexually active at baseline (n=331)	5.97**	.007
Subgroup: Respondent age		
Respondent less than age 15 (n=867)	14.04***	.000
Respondent age 15 or older (n=573)	5.77***	.000
Attitudes toward protection^d		
Subgroup: Respondent age		
Respondent less than age 15 (n=867)	0.13***	.000
Respondent age 15 or older (n=573)	0.05	.098
Motivation to delay childbearing^d		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,106)	-0.04	.294
Ever sexually active at baseline (n=329)	0.12	.077

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

Note: Impact estimates for subgroups are shown only if a test for differences in impacts among the subgroups met the study criterion for statistical significance ($p < .05$). For example, a test result indicated that the treatment effect on knowledge of pregnancy risk was significantly different for male and female youth.

^a This column shows the estimated treatment effect (treatment/control difference in percentage reporting engaging in the behavior) 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 from 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 E: SENSITIVITY ANALYSES ON HISPANIC-ONLY SAMPLE

Exhibit E.11: Longer-Term Effects on Non-Behavioral Intermediate Outcomes by Subgroup for Hispanic Sample Only

	Treatment Effect ^a	p-Value ^b
Knowledge of pregnancy risk^c		
Subgroup: Respondent gender		
Male (n=571)	-3.88	.123
Female (n=783)	6.25**	.003
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,042)	4.12*	.024
Ever sexually active at baseline (n=312)	-5.16	.132
Knowledge of STI risk^c		
Subgroup: Respondent gender		
Male (n=571)	2.99	.097
Female (n=783)	8.48***	.000
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,042)	7.44***	.000
Ever sexually active at baseline (n=312)	1.90	.437
Attitudes toward risky sexual behavior^d		
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,035)	0.10	.897
Ever sexually active at baseline (n=310)	-4.55**	.002
Intend to use condom if have sexual intercourse in next 12 months^e		
Subgroup: Respondent age		
Respondent less than age 15 (n=814)	-3.71~	.078
Respondent age 15 or older (n=528)	6.36*	.015
Subgroup: Respondent sexual experience at baseline		
Never sexually active at baseline (n=1,034)	-2.49	.177
Ever sexually active at baseline (n=308)	10.50**	.003

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

Note: Impact estimates for subgroups are shown only if a test for differences in impacts among the subgroups met the study criterion for statistical significance ($p < .05$). For example, a test result indicated that the treatment effect on knowledge of pregnancy risk was significantly different for male and female youth.

^a This column shows the estimated treatment effect (treatment/control difference in percentage reporting engaging in the behavior) 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 from zero.

^c Scores represent the average percentage of items answered correctly.

^d Score represents the average percentage of items agreed with.

^e Dichotomous variable, reported as percentage of respondents who responded affirmatively.

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

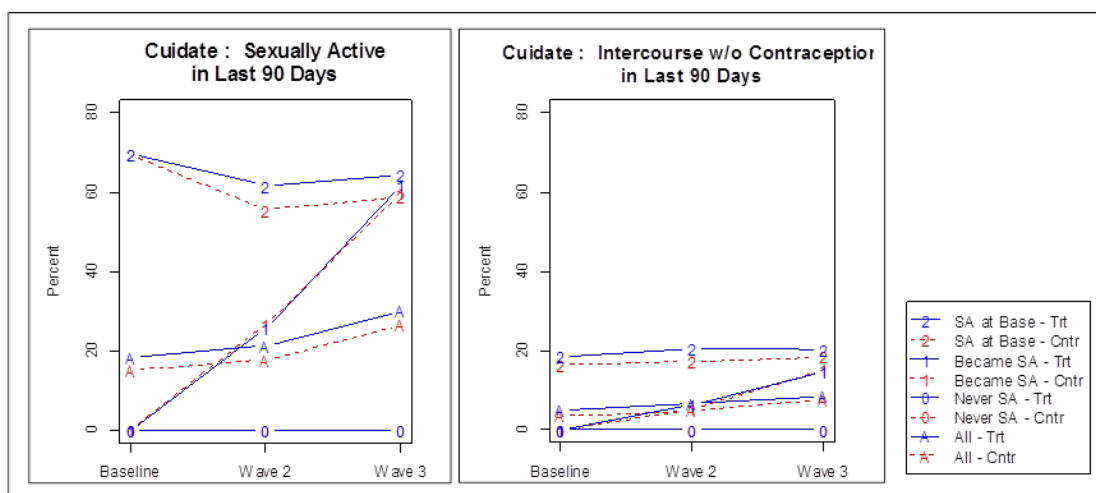
Appendix F: Preliminary Descriptive Findings on Subgroups of Youth Based on Sexual Activity Throughout the Study

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

In Exhibit F.1, we plot the confirmatory outcomes for the treatment and control groups at each of the study data collection time points for all study youth (plotting symbol=A), youth you 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 the graph, which plots sexual activity in the last 90 days, we see a potential difference in impacts between groups 1 and 2, such that the program may have an unintended impact on sexual activity over time for youth who were sexually active at baseline. There appears to be no difference between the treatment and control groups for youth who became sexually active throughout the study. These potential group differences are less apparent in the graph of recent sexual intercourse without contraception. In both graphs, the difference in trajectories for the two groups is also notable.

Exhibit F.1: Plots of Confirmatory 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 – 653/458, 6 months – 643/454, 18 months – 592/407; Group 1: Baseline – 250/175, 6 months – 252/176, 18 months – 254/172; Group 2: Baseline – 326/177, 6 months – 313/173, 18 months – 287/158. For sexual intercourse without contraception in last 90 days, treatment/control sample sizes are as follows: Group 0: Baseline – 653/458, 6 months – 643/454, 18 months – 592/407; Group 1: Baseline – 250/175, 6 months – 253/176, 18 months – 254/172; Group 2: Baseline – 328/177, 6 months – 313/173, 18 months – 287/158. Wave 2 = short-term follow-up (6 months after baseline); Wave 3 = longer-term follow-up (18 months after baseline).

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 between the subgroups of youth who were sexually active at baseline versus those who became sexually active during the study period. Further analyses of the differences between these groups would be exploratory and non-experimental, and thus are not included in this report.

Appendix G: Supporting Tables

Exhibit G.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	14.39	14.37	0.02	.629
Grade				
Mean	9.19	9.13	0.05*	.015
Gender				
Female ^c	52.98	52.98	0.00	n/a
Race/Ethnicity^d				
Hispanic	71.40	70.10	1.30	.506
Black	3.73	5.83	-2.10*	.027
White	17.96	18.49	-0.53	.749
Other	6.91	5.58	1.32	.231
Family structure and relationships				
Lives with biological parents	94.30	90.71	3.60**	.003
Feels very close to and cared for by father	44.45	46.06	-1.61	.499
Feels very close to and cared for by mother	59.13	60.29	-1.16	.611
Risk behaviors				
Ever smoked cigarettes	18.64	18.41	0.23	.897
Ever drank alcohol	46.06	48.08	-2.03	.355
Ever used marijuana	25.32	25.90	-0.58	.763
Sexual activity				
Ever sexually active ^e	25.22	21.99	3.23	.074
Recently sexually active (in last 90 days) ^e	17.38	14.69	2.69	.095
Sexual intercourse in the last 90 days	14.49	12.37	2.12	.164
Oral sex in the last 90 days	12.66	10.34	2.32	.107
Anal sex in the last 90 days ^e	4.25	3.42	0.84	.484
Sexual risk behavior				
Sexual intercourse without birth control in the last 90 days	4.38	3.48	0.90	.328
Sexual intercourse without a condom in the last 90 days	8.33	6.44	1.89	.112
Oral sex without a condom in the last 90 days	11.31	9.17	2.14	.119
Anal sex without a condom in the last 90 days ^e	2.77	2.05	0.72	.464
Knowledge^f				
Knowledge of pregnancy risk	48.37	47.98	0.39	.831
Knowledge of STI risk	38.61	39.36	-0.75	.556
Attitudes^g				
Attitudes toward protection	3.07	3.06	0.00	.830
Intentions				
Intentions to have sexual intercourse in the next 12 months	31.91	27.50	4.41*	.023
Intentions to have oral sex in the next 12 months	25.10	22.76	2.34	.205
Intentions to use birth control if they were to have sexual intercourse	92.07	91.99	0.08	.953
Intentions to use a condom if they were to have sexual intercourse	92.99	94.72	-1.73	.132

APPENDIX G: SUPPORTING TABLES

Source: Baseline survey administered prior to randomization.

Note: Results in this table are based on the analytic sample of 1,143–2,022 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 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 it 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 La Alianza and Community Action Partnership, *sexual activity* refers to sexual intercourse, oral sex, and/or anal sex. Youth were not asked about anal sex in Touchstone.

^f Knowledge variables are composite scale scores representing the percentage of items answered correctly.

^g Attitudes variable is a composite scale score, with higher scores indicating more positive attitudes.

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

APPENDIX G: SUPPORTING TABLES

Exhibit G.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	14.38	14.37	0.01	.807
Grade				
Mean	9.18	9.14	0.04	.053
Gender^c				
Female	54.03	54.03	0.00	n/a
Race/Ethnicity^d				
Hispanic	72.58	69.62	2.96	.139
Black	3.65	5.65	-1.99*	.043
White	16.75	18.68	-1.93	.246
Other	7.01	6.05	0.97	.407
Family structure and relationships				
Lives with biological parents	94.37	90.91	3.46**	.006
Feels very close to and cared for by father	44.14	46.96	-2.82	.251
Feels very close to and cared for by mother	58.81	60.73	-1.93	.414
Risk behaviors				
Ever smoked cigarettes	17.46	17.39	0.08	.966
Ever drank alcohol	45.87	47.71	-1.83	.423
Ever used marijuana	24.52	26.22	-1.70	.395
Sexual activity				
Ever sexually active ^e	23.90	22.01	1.90	.310
Recently sexually active (in last 90 days) ^e	16.23	14.61	1.62	.326
Sexual intercourse in the last 90 days	13.47	12.36	1.11	.475
Oral sex in the last 90 days	11.86	9.97	1.89	.204
Anal sex in the last 90 days ^e	3.54	3.19	0.35	.769
Sexual risk behavior				
Sexual intercourse without birth control in the last 90 days	4.38	2.95	1.43	.131
Sexual intercourse without a condom in the last 90 days	7.86	6.04	1.82	.134
Oral sex without a condom in the last 90 days	10.70	8.57	2.13	.131
Anal sex without a condom in the last 90 days ^e	2.64	1.72	0.92	.343
Consequences of sexual risk behavior				
Ever pregnant or gotten someone pregnant (lifetime)	1.70	1.81	-0.11	.867
Diagnosed with STI in the last 12 months	0.43	0.42	0.01	.974
Knowledge^f				
Knowledge of pregnancy risk	48.45	47.81	0.64	.737
Knowledge of STI risk	39.26	39.36	-0.11	.936
Attitudes^g				
Attitudes toward protection	3.08	3.07	0.01	.653
Intentions				
Intentions to have sexual intercourse in the next 12 months	30.46	27.25	3.21	.112
Intentions to have oral sex in the next 12 months	24.23	22.38	1.85	.334
Intentions to use birth control if they were to have sexual intercourse	92.67	92.58	0.09	.945
Intentions to use a condom if they were to have sexual intercourse	93.22	95.22	-2.00	.085

APPENDIX G: SUPPORTING TABLES

Source: Baseline survey administered prior to randomization.

Note: Results in this table are based on the analytic sample of 1,067–,885 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 it 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 La Alianza and Community Action Partnership, *sexual activity* refers to sexual intercourse, oral sex, and anal sex. Youth were not asked about anal sex in Touchstone.

^f Knowledge variables are composite scale scores representing the percentage of items answered correctly.

^g Attitudes variable is a composite scale score, with higher scores indicating more positive attitudes.

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