



Evaluation of Adolescent
Pregnancy Prevention Approaches

OAH Evaluation Report

Impact Report from the Evaluation of Adolescent Pregnancy Prevention Approaches



Final Impacts of Teen PEP (Teen Prevention Education Program) in New Jersey and North Carolina High Schools

September 2016



Purpose statement: This study reports final findings from a large-scale demonstration project and evaluation of the Teen Prevention Education Program (Teen PEP), as implemented in New Jersey and North Carolina high schools. Teen PEP is an in-school, peer-to-peer sexual health promotion program that combines peer-led interactive workshops and peer-driven school-wide initiatives in an effort to reduce sexual risk behaviors and associated outcomes among high school students. The study reports final impacts of the program on sexual risk behaviors measured about 18 months after the program's conclusion.

September 2016

Recommended citation:

U.S. Department of Health and Human Services, Office of Adolescent Health, *Final Impacts of Teen PEP (Teen Prevention Education Program) in New Jersey and North Carolina High Schools*. Washington, D.C., September 2016.

This document is available at <http://www.hhs.gov/ash/oah/>

OAH is grateful to the authors Dana Rotz, Brian Goesling, Molly Crofton, Jennifer Manlove, and Kate Welti

CONTENTS

A. About the Teen PEP program	2
B. Study design, data, and analysis.....	4
1. School recruitment	4
2. Student enrollment and retention	6
3. Intervention and comparison conditions	6
4. Data and measures.....	7
5. Analysis methods	8
C. Characteristics of study participants.....	9
D. Impacts of Teen PEP on sexual risk behaviors.....	13
1. Prevalence of sexual activity.....	13
2. Prevalence of unprotected sex	15
E. Discussion and conclusion	17
REFERENCES.....	21
APPENDIX A DATA AND MEASURES	A-1
APPENDIX B ANALYTIC METHODS.....	B-1
APPENDIX C ROBUSTNESS ANALYSIS.....	C-1
APPENDIX D STATE-SPECIFIC ESTIMATES	D-1

TABLES

1.	Baseline characteristics for full sample.....	10
2.	Baseline characteristics for propensity-score matched sample.....	12
3.	Impacts of Teen PEP on sexual activity, estimated by regression analysis.....	15
4.	Impacts of Teen PEP on unprotected sex, estimated by regression analysis.....	17
A.1.	Baseline characteristics of study sample.....	A-5
B.1.	Candidate and selected variables for propensity-score regression.....	B-4
C.1.	Impacts of Teen PEP on outcomes, regression only controls for linear terms from propensity score regression.....	C-4
C.2.	Impacts of Teen PEP on outcomes, logit model used for all outcomes.....	C-5
C.3.	Impacts of Teen PEP on outcomes, estimate propensity score models by state.....	C-6
C.4.	Impacts of Teen PEP on outcomes, not adjusting for clustering.....	C-7
C.5.	Impacts of Teen PEP on outcomes, alternative coding of sexual behaviors 1.....	C-8
C.6.	Impacts of Teen PEP on outcomes, alternative coding of sexual behaviors 2.....	C-8
C.7.	Impacts of Teen PEP on outcomes, alternative coding of sexual behaviors 3.....	C-9
D.1.	Impacts of Teen PEP on sexual activity estimated by regression analysis, North Carolina sample.....	D-6
D.2.	Impacts of Teen PEP on unprotected sex estimated by regression analysis, North Carolina sample.....	D-8
D.3.	Impacts of Teen PEP on sexual activity estimated by regression analysis, New Jersey sample.....	D-11
D.4.	Impacts of Teen PEP on unprotected sex estimated by regression analysis, New Jersey sample.....	D-13

FIGURES

1. Flow of study schools and students 5

2. Impacts of Teen PEP on sexual activity, difference-in-differences in full sample..... 14

3. Impacts of Teen PEP on unprotected sex, difference-in-differences in full sample 16

D.1. Impacts of Teen PEP on sexual activity, difference-in-differences in full North Carolina sample..... D-5

D.2. Impacts of Teen PEP on unprotected sex, difference-in-differences in full North Carolina sample..... D-7

D.3. Impacts of Teen PEP on sexual activity, difference-in-differences in full New Jersey sample..... D-10

D.4. Impacts of Teen PEP on unprotected sex, difference-in-differences in full New Jersey sample..... D-12

This page has been left blank for double-sided copying.

ACKNOWLEDGEMENTS

This evaluation was made possible only through the hard work and support of many people and organizations. First, we wish to acknowledge the support of staff from the Office of Adolescent Health (OAH) in the U.S. Department of Health and Human Services (HHS). We especially thank Amy Farb, the OAH project officer for the Evaluation of Adolescent Pregnancy Prevention Approaches (PPA) study, who has provided outstanding support and guidance throughout the project, and Alexandra Warner of OAH, who provided valuable support in preparing the report. We also thank staff from the Office of Planning, Research and Evaluation within the Administration for Children and Families, especially Seth Chamberlain, for their support and guidance during the early site recruitment and design phase of the evaluation.

Special thanks goes to staff from the Center for Supportive Schools (CSS) and HiTops, Inc. for their tireless efforts recruiting and retaining study schools and supporting program implementation and study data collection. We especially thank Sherry Barr of CSS, for her unparalleled commitment to and enthusiasm for the program and study. We also thank staff from Abt Associates, in particular Carolyn Layzer, for leading the study design and data collection efforts in North Carolina high schools.

At Mathematica Policy Research, we thank Melissa Thomas for her long-standing commitment to the study and for successfully leading a highly complex, extended study data collection effort; Christopher Trenholm for his leadership during earlier phases of this research; and Lara Hulsey for valuable comments on this report. We also thank Jennifer Littel for editorial assistance and Colleen Fitts for producing the report. At Child Trends, we thank Heather Fish for helpful research assistance.

Finally, we extend our greatest thanks and gratitude to the youth, teachers, and school and district staff who agreed to participate in the study. We hope that our report does justice to the time and effort they devoted to making the study possible.

Dana Rotz

Brian Goesling

Molly Crofton

Mathematica Policy Research

Jennifer Manlove

Kate Welti

Child Trends

This page has been left blank for double-sided copying.

This report presents final impact findings from a large-scale demonstration project and evaluation of the Teen Prevention Education Program (Teen PEP), an in-school, peer-to-peer sexual health promotion program that aims to reduce sexual risk behaviors and associated outcomes among high school students. Although teen pregnancy rates in the United States have decreased over the past 25 years (Martin et al. 2015), teenage pregnancy remains a serious concern. Teen parenting is linked to negative outcomes for both teen parents and their children. Teen mothers are less likely to graduate high school, have lower earnings, receive public assistance for longer periods, and are more likely to be single parents (Hoffman 2008; Perper et al. 2010). Children of teen mothers have worse educational, criminal, and health outcomes as well (Hoffman 2008). Furthermore, teen pregnancy is associated with high costs to the general public; in 2010, teen pregnancy and childbirth in the United States cost taxpayers more than \$9.4 billion (National Campaign to Prevent Teen and Unplanned Pregnancy 2013). Teen PEP, developed and implemented by the Center for Supportive Schools (CSS) and HiTops, Inc., seeks to reduce rates of teen pregnancy and associated sexual risk behaviors through a combination of school-based, peer-led interactive workshops and school-wide initiatives.

In an earlier report (Rotz et al. 2016), we examined the short-term impacts of Teen PEP on both student sexual risk behaviors and intermediate outcomes such as student knowledge, attitudes, and exposure to information on reproductive health topics. Drawing on data from a volunteer sample of New Jersey and North Carolina high schools, we found that students in schools that offered Teen PEP reported greater exposure to information on reproductive health topics such as abstinence, birth control methods, and sexually transmitted infections (STIs) than did students in schools that did not offer the program. Students in schools that offered Teen PEP also reported higher average scores on an index measuring student knowledge of preventing pregnancy and STI/HIV transmission. However, we found no evidence that these impacts on intermediate student outcomes led to decreases in the incidence of sexual activity or unprotected sex. Students in schools that offered Teen PEP were just as likely as students in other schools to report having had sex or unprotected sex.

In the present report, we extend these results by examining the program's longer-term impacts on student sexual risk behaviors. Teen PEP is based on a theory of peer influence and school-wide cultural change, meaning that changes in student behaviors are likely to emerge only as students interact with and are influenced by their peers. Thus, it may take time for the program to impact student behavior. Our earlier interim report drew on survey data collected from students in fall of 10th grade, about 6 to 12 months after students had received the Teen PEP workshops as 9th graders. In this report, we present results from a longer-term survey, conducted about a year later, when students were in 11th grade. Using data from this survey, we examine the longer-term impacts of Teen PEP on student sexual risk behaviors, measured 18 to 24 months after students had received the 9th grade Teen PEP workshops.

This evaluation has involved a unique partnership and collaboration among several organizations. In 2009, Teen PEP was selected as one of the first sites in the Evaluation of Adolescent Pregnancy Prevention Approaches (PPA), a major federal effort to expand available evidence on effective ways to prevent and reduce pregnancy and related sexual risk behaviors among teens in the United States. Mathematica Policy Research and its partners, Child Trends and Twin Peaks Partners, LLC, under contract with the Office of Adolescent Health (OAH) within the U.S. Department of Health and Human Services, are conducting the evaluation. The

original PPA study design called for evaluating the effects of Teen PEP among high schools in New Jersey. In 2010, CSS received a separate competitive grant from OAH to expand and evaluate Teen PEP in another state, North Carolina. The evaluation of the North Carolina program was originally designed by researchers from Abt Associates. Under the guidance of OAH, these planned evaluation activities were ultimately combined into a single, unified evaluation of Teen PEP across two states: New Jersey and North Carolina. Researchers from the PPA study team led the combined impact study and conducted an implementation study of Teen PEP in New Jersey high schools. Researchers from Abt Associates led an implementation study of Teen PEP in North Carolina high schools and supported data collection in that state for the combined impact study.

The remainder of this report comprises five sections. We begin by providing a detailed description of Teen PEP. We then describe our evaluation's design, the data and measures we used, and our analysis methods. Next, we describe the characteristics of the students included in the analysis. We then present the final impact findings for the measures of student sexual risk behaviors. We conclude by discussing the implications of these results. We provide additional technical documentation in Appendices A and B and supplementary analyses in Appendices C and D of this report.

A. About the Teen PEP program

CSS and HiTops have a long history of educating students on the dangers of risky behavior. In the 1990s, the organizations developed a youth HIV/AIDS prevention program at the request of the New Jersey Department of Health. As a result of early implementation efforts, staff at the organizations became aware of the need for a more sustained and comprehensive school-based sexual education program. They thus developed Teen PEP. By 2009, the program was well established in more than 50 public high schools throughout New Jersey. Starting in 2005, the program also expanded to a small number of North Carolina schools, in partnership with the North Carolina Department of Health and Human Services, Teen Pregnancy Prevention Initiatives and the North Carolina Department of Public Instruction, Healthy Schools. In both states, school stakeholders work with CSS staff to implement the program.

As currently implemented, Teen PEP's comprehensive sexual education curriculum is interactive and dynamic and aims to build strong connections among participating students, staff, and the school community. In this section, we summarize the main components of the program. For further details on the program model and its implementation, see Asheer et al. (2014); Layzer et al. (2014); Layzer and Rosapep (2012, 2013); and Rotz et al. (2016).

At each school implementing Teen PEP, a group of 11th- and 12th-grade students with identified leadership potential is selected through an application process to serve as peer educators and trained by faculty advisors to become leaders and role models. Peer educators begin their training with a mandatory retreat, designed to foster trust and cohesion between the students and faculty advisors. They then participate in either a 45-minute class each school day throughout the academic year or a 90-minute class each school day for one semester, usually as an elective course or as a replacement course for health or physical education. The course is designed to prepare students to deliver outreach workshops on sexual health topics to 9th-grade

students, their parents, and other family members. Faculty advisors deliver 10 core units to the peer educators, incorporating experiential and activity-based learning.

The Teen PEP curriculum describes the five core workshops that peer educators deliver to 9th-grade participants. The 90-minute workshops focus on topics most relevant to reducing risky behaviors:

1. **Let's Wait Awhile: Postponing Sexual Involvement.** Peer educators and workshop participants describe reasons why teens do and do not become sexually involved, and possible consequences of early sexual involvement; identify relationship qualities that are important to have before beginning a sexual relationship; and demonstrate negotiation and refusal skills.
1. **Later, Baby: Pregnancy Prevention.** Peer educators and workshop participants identify behaviors that put teens at risk for unintended pregnancy; identify solutions to barriers that get in the way of teens using condoms, practicing birth control, or seeking guidance at a family-planning clinic; describe at least three methods for preventing pregnancy; and identify the location of a nearby family-planning clinic.
2. **Don't Pass It On: Preventing Sexually Transmitted Infections.** Peer educators and workshop participants describe the identification, symptoms, treatment, and long-term consequences of the most common STIs among teens; demonstrate a greater understanding of how STIs are spread; and identify personal strategies for preventing the spread of STIs.
3. **Break the Silence: HIV/AIDS Prevention.** Peer educators and workshop participants describe the two most common ways teens become infected with HIV/AIDS, identify behaviors that will decrease the risk of HIV infection, name strategies for reducing the risk of contracting HIV/AIDS, describe the steps to using a condom correctly, and increase motivation for using risk-reduction strategies.
4. **Sex on the Rocks: Alcohol, Other Drugs, and Sexual Decision Making.** Peer educators and workshop participants identify steps to decision making and the consequences of making sexual decisions under the influence of alcohol and other drugs. They also demonstrate refusal skills to resist the pressure to use alcohol and other drugs.

The peer educators also deliver a sixth workshop (**Talk to Me: A Family Night**) for parents/guardians, family members, and the broader school community. This workshop is designed to help parents or caregivers identify their personal attitudes and values regarding sexuality, become more comfortable talking about sex and sexuality with teens, and develop their understanding of how to initiate conversations about sex and sexuality with teens.

Both the classes that the peer educators take and the workshops they conduct are designed to engage and appeal to teens. In workshops, peer educators use accessible, plain language and humor to convey main points and messages as part of the activities. Each workshop consists of skits, skill-building activities, and small-group discussions that the peer educators facilitate. Activities incorporate and emphasize communication with peers and parents, problem solving, decision making, and negotiation and refusal skills.

In addition to the workshops and family night, peer educators also lead a school-wide campaign to promote positive cultural change. This campaign reinforces workshop messages for participating 9th-grade students and can help spread the Teen PEP messages to the broader school community. The Teen PEP curriculum does not prescribe the form that the campaign should take; students are asked to shape the campaign based on their own experiences of what activities are most influential and memorable.

B. Study design, data, and analysis

We conducted our evaluation of Teen PEP with a volunteer sample of New Jersey and North Carolina high schools. Among the schools recruited for the study, we randomly assigned just over half to an “early” implementation group that could begin implementing Teen PEP the first year after enrolling in the study (the intervention group). We randomly assigned the other schools to a “later” implementation group that had to delay implementing Teen PEP for at least 1.5 years (the comparison group). To assess the impacts of Teen PEP, we administered self-reported surveys to 9th-grade students in both the intervention and comparison schools at three points: (1) a baseline survey administered in 9th grade, before the start of the Teen PEP workshops in the intervention schools; (2) an interim follow-up survey administered in late fall or winter of 10th grade, about 6 to 12 months after the workshops had ended; and (3) a final follow-up survey administered in late fall or winter of 11th grade, about 18 to 24 months after the workshops had ended. For the present report, we focus on data from the final follow-up survey.

1. School recruitment

We recruited an initial sample of 30 schools in New Jersey and North Carolina, planning to implement Teen PEP over the course of three school years: 2011–2012, 2012–2013, and 2013–2014 (Figure 1). We began the recruitment effort in New Jersey, seeking schools interested in implementing Teen PEP but not already doing so. We extended our recruitment efforts to North Carolina after CSS received a federal grant to support a demonstration project in schools outside of New Jersey. The North Carolina schools participating in the evaluation were some of the first in the state to receive the program, whereas more than 50 schools in New Jersey implemented Teen PEP before the evaluation. In total, we recruited 13 North Carolina schools and 17 New Jersey schools for the evaluation.

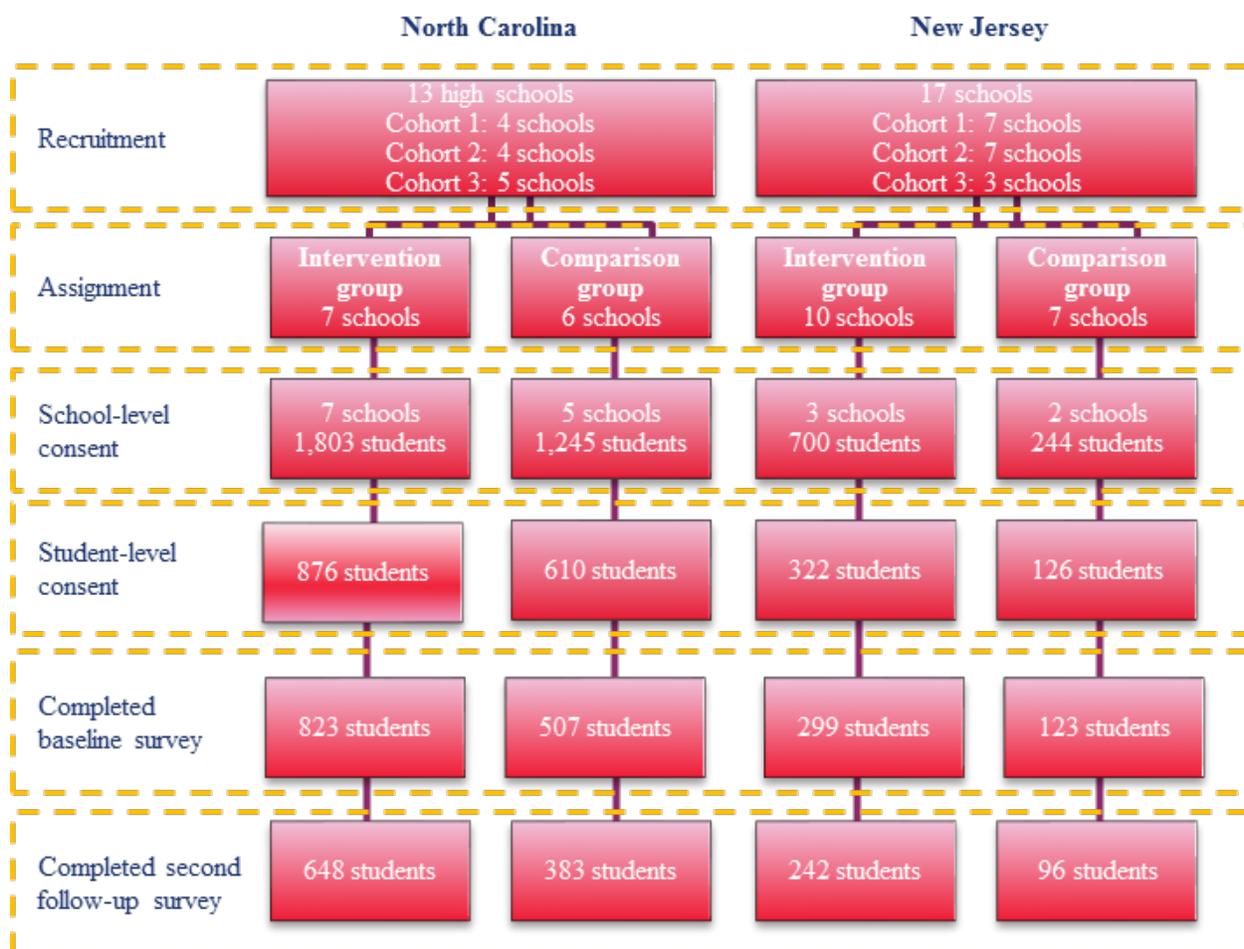
Within each cohort and state, we randomly assigned schools to the intervention and comparison conditions. In New Jersey, we further grouped schools into pairs or triplets based on school characteristics before random assignment in an effort to increase balance between study groups. In practice, this grouping may not have improved balance to a large degree, because each New Jersey cohort included only a small number of schools, with some schools being notably different from others in the cohort. We did not group schools in North Carolina in this way.

A large number of schools ultimately dropped out of the study after random assignment but before data collection began. Attrition was especially pronounced in the New Jersey sample. Of the 17 high schools initially recruited in New Jersey, 6 schools (35 percent) dropped out for various reasons, including concerns about the study survey, school closure, and a district-wide policy against evaluation. Teen PEP requires a planning period of several months, necessitating a lengthy time between school recruitment and other study activities, which might have

exacerbated school concerns about participating in the evaluation and facilitated attrition. The evaluation also lost another 6 schools (35 percent) from the New Jersey sample because these had been randomized as part of the same matched pair or triplet as the schools that dropped out. Of the 13 high schools initially recruited in North Carolina, 1 school (8 percent) left the sample before any data collection. We provide a more detailed description of the reasons for school attrition in our earlier interim report (Rotz et al. 2016).

For the current analysis, we used data from the 17 high schools that remained in the study throughout the baseline and follow-up data collection. This sample includes 12 of the original 13 high schools from the North Carolina sample (92 percent) and 5 of the original 17 high schools from the New Jersey sample (29 percent). In accordance with our initial design for the study, we conduct our impact analysis by combining the data across the two states. However, we also examined the data for the two states separately in additional exploratory analyses (see Appendix D).

Figure 1. Flow of study schools and students



Source: Teen PEP study information system.

Note: Cohort 1 implemented Teen PEP in school year 2011–2012, Cohort 2 in 2012–2013, and Cohort 3 in 2013–2014. All numbers are cumulative; the number of students in the final row includes only students who also completed the baseline survey.

2. Student enrollment and retention

In both intervention and comparison schools, we recruited 9th-grade students for the evaluation. Students had to obtain the written permission of a parent or guardian before the baseline survey to participate in the study data collection. The study team offered individuals a \$5 gift card for returning a completed consent form, regardless of whether their parent or guardian provided consent. In some schools (where allowed by the school stakeholder team), verbal consent was permitted as well. In these cases, the study team contacted parents and guardians of those who did not complete the written consent form via telephone and read the form aloud. The study team members then completed the form as indicated by the parent or guardian and sent a copy to the student's home. In intervention schools, participation in the study or study surveys did not impact whether a student could receive the program.

This consent process resulted in about 48 percent of all eligible students receiving permission to participate in the study. Just over two-thirds of students returned a consent form, and the parents of about three-quarters of students who returned a form consented to participate in the evaluation. Consent rates were very similar for individuals in intervention and comparison groups taken all together but varied across schools, from 37 to 67 percent. We provide more detailed information on the school-level consent rates in our earlier interim report (Rotz et al. 2016).

Among those students who received permission to participation in the study, participation in the study surveys was reasonably high. In the sample of students who returned an affirmative consent form, 94 percent of the intervention group and 86 percent of the comparison group completed the baseline survey. Rates of response to the final follow-up survey were more modest, with 78 percent of the intervention group and 72 percent of the comparison group completing this survey. As with consent, these rates also varied widely by school, from 61 to 99 percent for the baseline survey and from 57 to 89 percent for the final follow-up survey.

3. Intervention and comparison conditions

With the support of CSS and HiTops, all schools in the intervention group implemented Teen PEP during the prescribed period, one school year for New Jersey schools and one semester for North Carolina schools (which used block scheduling). CSS and HiTops provided extensive support to intervention schools to facilitate the successful implementation of Teen PEP. Both organizations trained faculty advisors in what to look for when recruiting and selecting the peer educators who lead Teen PEP, how to prepare peer educators to conduct workshops, and how to observe the quality of the workshops. After implementation had begun, CSS and HiTops also offered additional training and technical assistance to schools that were delivering Teen PEP. Teen PEP staff monitored program delivery by attending and observing peer educator classes, peer-led workshops, and stakeholder team meetings. Staff also frequently offered written and oral feedback to ensure schools' adherence to the program model.

Findings from our accompanying implementation studies indicate that schools implemented most key components of Teen PEP as intended, although some schools struggled with certain aspects of the program (Asheer et al. 2014; Layzer and Rosapep 2012, 2013). For example, some faculty advisors reported that delivering the activity-based learning model to peer educators was difficult, an issue that was compounded by many faculty members' lack of experience with

classroom teaching and facilitating intensive, structured programs such as Teen PEP. Also, some schools struggled with the logistical demands of providing the small-group workshops to as many as 300 9th-grade students, which often required that peer educators deliver the same content multiple times. Although most of the study schools implemented all of the scheduled Teen PEP workshops, in some cases attendance was low. For example, in one North Carolina school, only 62 percent of students attended four or more of the six Teen PEP workshops (including the family night, see Asheer et al. 2014). Finally, finding classroom space for workshops impeded program operations in some cases. Limited buy-in and involvement of the stakeholder team intensified these issues in some schools.

During the intervention period, schools in the comparison group could implement any existing sexual health programs or curricula other than Teen PEP. Such programs varied across schools but especially across states. Students in most New Jersey schools take a health class that includes comprehensive sexual education. In contrast, few comparison schools in the North Carolina sample had exposure to comprehensive sexual education during their health class, despite a 2009 state requirement for schools to provide sexual education. Instead, students in North Carolina schools are more commonly exposed to abstinence-until-marriage curricula. For example, survey data collected from high school principals in North Carolina in 2010 indicate that only 21 percent of high schools had curricula that exposed students to condom-use topics within a required course (Mitchell and Greene 2011). The same data also show that most North Carolina health teachers had not received professional development covering HIV- or pregnancy-prevention information in the past two years, suggesting that educators might not have access to the most up-to-date information on these topics. Additionally, access to reproductive health services varied greatly by school. One school had a collocated clinic providing free health services, but another was more than 50 miles from such resources.

4. Data and measures

For the present report, we draw on data from two of the three rounds of survey data collection: (1) the baseline survey administered in late fall or early winter of 9th grade, before the delivery of the first Teen PEP workshop in the intervention schools and (2) the final follow-up survey administered about 18 to 24 months after the workshops, in the fall of the students' 11th-grade year. Students received gift-card incentives for completing the follow-up survey (a \$10 incentive if the student completed the survey at the time it was administered at the school and a \$25 incentive if the student completed the survey over the phone). Our interim report presents comparable impact findings for the data collected from an earlier follow-up survey (Rotz et al. 2016).

To assess the longer-term impacts of Teen PEP on student sexual risk behaviors, we used the final follow-up survey data and constructed five key outcome measures. Three of these outcomes assess the **prevalence of sexual activity**: (1) a binary indicator equal to one if a student reported having sexual intercourse in the past three months (and zero otherwise), (2) a binary indicator equal to one if a student reported ever having sexual intercourse (and zero otherwise), and (3) a count measure of the number of lifetime sexual partners the student reported. The other two outcomes measure the **prevalence of unprotected sex**: (1) a binary indicator equal to one if a student reported having sexual intercourse without a condom in the past three months (and zero otherwise), and (2) a binary indicator equal to one if a student

reported having sexual intercourse without any effective birth control method in the past three months (and zero otherwise).

Before analysis, we specified that the indicators for having had intercourse within the past three months and having had intercourse without a condom within the past three months would serve as confirmatory outcomes. We chose student reports of sexual activity in the past three months as our confirmatory measure for the prevalence of sexual activity because it captures both primary and secondary abstinence. We chose student reports of sexual activity without a condom in the past three months as our confirmatory measure for the prevalence of unprotected sex because of Teen PEP's focus on both HIV/STI-transmission prevention and pregnancy prevention. Using a broader measure of contraceptive use as a confirmatory outcome would capture only pregnancy prevention. We provide additional details on each outcome in Appendix A.

5. Analysis methods

We designed our analysis methods accounting for two key issues. First, as discussed earlier in this section, a relatively large number of schools dropped out of the study after random assignment, especially from the New Jersey sample. Because schools left the study in a non-random manner, this school-level attrition has the potential to compromise the validity of the original random assignment evaluation design, and we cannot be confident that any differences between the study groups in our analytic sample at baseline are due only to chance. Second, as we show in detail below, among the 17 schools that were retained for the study, the baseline characteristics of the student samples differed markedly between the intervention and comparison groups. For example, students in intervention schools were about half as likely as students in comparison schools to report having had sex without using a condom in the three months before the baseline survey. We must adjust for these pre-existing differences when analyzing data from the follow-up survey to avoid making false claims about the program's impacts on student outcomes.

We used four different analytic techniques to mitigate the concern that attrition and baseline differences will bias our estimates of the impact of Teen PEP. First, we conducted a difference-in-differences (DD) analysis, comparing changes in outcomes over time across intervention and comparison schools. Next, we used ordinary least squares regression to estimate program impacts, controlling for both baseline measures of outcomes and other predictors of behavior. Additionally, we used two different quasi-experimental (comparison group) approaches that restricted the study sample to better-matched subsets of the original sample: propensity-score trimming and propensity-score matching. Propensity-score trimming uses a propensity-score model to restrict the sample to students who are not very different from the average student in the intervention or comparison group. This can reduce baseline differences between the groups. Propensity-score matching directly selects a comparison group that is more similar to the existing intervention group. In both cases, we used ordinary least squares to estimate impacts of Teen PEP within the reduced samples. We provide a more detailed description of each technique in Appendix B.

We chose the propensity-score matching approach as our primary estimation method because past research suggests it is the most likely of these four techniques to provide accurate impacts (Funk et al. 2011; Dehejia and Wahba 1999; Imbens and Wooldridge 2009). We present

estimates using all four methods to demonstrate the robustness of our results to alternative analytic decisions, and to show whether and how the propensity-score matching affects the direction and magnitude of our impact estimates.

C. Characteristics of study participants

The baseline characteristics of the full study sample reflect the characteristics of the schools recruited for the study (Table 1). The average age of students was just over 15 years old, the sample was split roughly evenly between female and male students, and the students had diverse racial/ethnic backgrounds. Just over 8 in 10 students listed English as the main language spoken in their homes. Rates of lifetime cigarette use, alcohol use, marijuana use, and sexual activity are similar to national averages for 9th-grade students reported in the Youth Risk Behavior Survey (Kann et al. 2014).

Students typically reported some previous exposure to the topics that the Teen PEP curriculum covers (Table 1). Within the year before our baseline survey, more than three-quarters of students in both groups had received information related to relationships, STIs, saying no to sex, and how babies are made; about two-thirds had received information on abstinence. The share of students reporting learning about methods of birth control, where to get birth control, and talking to a partner about sex were all somewhat lower; around half of all students reported receiving this information.

Students in the intervention and comparison groups differed on many key characteristics within the full sample (Table 1). Students in intervention schools were 0.3 years older than those in comparison schools, a statistically significant difference ($p = 0.01$). Racial composition also varied between intervention and comparison schools. In intervention schools, 30 percent of students were Hispanic, 25 percent were black non-Hispanic, and 32 percent were white non-Hispanic. In comparison schools, the composition was 23 percent Hispanic, 43 percent black non-Hispanic, and 25 percent white non-Hispanic. The difference in share black is large (19 percentage points) and marginally statistically significant ($p = 0.08$). Students in intervention schools were more likely to live with their biological mother and to live with their biological father, with both differences statistically significant ($p = 0.04$). Nonsexual risk behaviors were also more common for comparison-group students than for intervention-group students. Students in the comparison group were more likely to have reported ever having smoked cigarettes ($p = 0.10$), having drunken alcohol in the past 30 days ($p = 0.06$), or having smoked marijuana in the past 30 days ($p = 0.04$).

Table 1. Baseline characteristics for full sample

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Sociodemographic characteristics				
Average age	15.0	15.3	-0.3**	0.01
Female	55.4	56.8	-1.4	0.59
Race				
Hispanic	30.0	22.6	7.4	0.50
White, non-Hispanic	32.3	25.2	7.1	0.54
Black, non-Hispanic	24.7	43.3	-18.6	0.08
Other race/ethnicity	12.8	7.7	5.1	0.09
Race missing	0.2	1.2	-1.0*	0.04
Main language spoken at home is English	81.5	85.2	-3.7	0.66
Biological mother living in home or main home	88.2	82.4	5.8*	0.04
Biological father living in home or main home	55.1	45.5	9.6*	0.04
Biological parents currently married	50.7	40.5	10.2	0.05
Ever smoked cigarettes	24.5	31.8	-7.3	0.10
Drank alcohol in the past 30 days	24.6	31.0	-6.5	0.06
Smoked marijuana in the past 30 days	13.9	20.6	-6.7*	0.04
Exposure to Teen PEP messages				
Received any information in past 12 months on:				
Relationships	86.6	85.3	1.3	0.52
Abstinence	67.8	64.2	3.6	0.31
Methods of birth control	49.7	48.8	0.9	0.80
Where to get birth control	43.3	42.2	1.1	0.78
STIs	80.7	75.3	5.4	0.09
Talking about sex with your partner	50.9	52.9	-2.0	0.67
Saying no to sex	73.9	73.5	0.4	0.93
How babies are made	85.4	79.5	5.9	0.06
Sexual behaviors				
Ever had sexual intercourse	21.9	31.1	-9.2	0.06
Number of lifetime sexual partners	0.69	0.91	-0.22	0.35
In the three months before survey				
Had sexual intercourse ^a	12.0	20.3	-8.3*	0.02
Had sexual intercourse without a condom ^a	7.4	15.3	-7.9*	0.02
Had sexual intercourse without any effective method of birth control	6.1	13.1	-7.0*	0.01
Ever had nonpenetrative sex	48.2	60.1	-11.9	0.06
Ever kissed member of opposite sex	81.8	89.3	-7.6*	0.03
Sample size	890	479		

Source: Teen PEP baseline survey.

Note: Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each school equal weight. *P*-values corrected for clustering at the school level.

^aConfirmatory outcome.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

Students in the full-sample intervention and comparison groups further exhibited different rates of baseline sexual activity (Table 1). In the three months before the baseline survey, students in comparison schools were 8 percentage points more likely to have had sex, 8 percentage points more likely to have had sex without a condom, and 7 percentage points more likely to have had sex without any effective method of birth control. All three differences are statistically significant ($p < 0.02$). The differences are particularly large when compared to mean rates of sexual activity in intervention schools; they imply that students in comparison schools were about twice as likely as students in intervention schools to engage in sexual risk behaviors. Additionally, only 22 percent of intervention-group students had ever had sex, compared to 31 percent of comparison-group students, a marginally statistically significant difference of 9 percentage points ($p = 0.06$). Individuals in the comparison group were also less likely to report having engaged in nonpenetrative sex (a marginally significant difference of 12 percentage points, $p = 0.06$) or having kissed a member of the opposite sex (a significant difference of 8 percentage points, $p = 0.03$).

Propensity-score matching decreased many of the cross-group differences in background characteristics (Table 2). The difference in the proportion of intervention- and comparison-group students who are black shrinks from 19 to 9 percentage points after propensity-score matching. Furthermore, in the propensity-score matched sample, there is little difference in the average age of students across study groups. Differences in parental marital status, students' living situations, and students' nonsexual risk behaviors also generally shrink with the matching. Furthermore, differences in prior exposure to the topics covered by Teen PEP remain small and statistically insignificant.

Propensity-score matching also greatly decreases the differences in most baseline measures of sexual risk behavior between students attending Teen PEP and comparison schools (Table 2). The differences in sexual initiation, recent sexual activity, and recent unprotected sex shrink to 2 percentage points and become statistically insignificant ($p > 0.15$). Differences in non-intercourse sexual behavior (nonpenetrative sex and kissing) also decrease to one percentage point or less ($p > 0.88$).

As these gaps shrink, one notable difference does emerge. In the propensity-score matched sample, the average intervention-group student reported 0.73 lifetime sexual partners, compared to 0.40 partners for comparison-group students (a difference of 0.33, $p = 0.04$). This difference is driven mainly by a small number of students in intervention schools reporting a large number of partners (not shown). Additionally, after propensity-score matching, a sizable and marginally significant difference remains in the share of students who reported drinking alcohol in the past 30 days. After matching, intervention-group students were 8 percentage points more likely than comparison-group students to report this behavior ($p = 0.06$). Despite these differences in both lifetime sexual partners and past alcohol consumption, the overall balance in baseline characteristics between the intervention and comparison groups is better in the propensity-score matched sample than in the full study sample. In addition, we control for measures of both lifetime sexual partners and past alcohol consumption at baseline in our regression analyses to mitigate concerns that these differences lead to biased results.

Table 2. Baseline characteristics for propensity-score matched sample

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Sociodemographic characteristics				
Average age	15.1	15.0	0.0	0.86
Female	56.8	53.9	2.8	0.49
Race				
Hispanic	30.1	24.3	5.7	0.64
White, non-Hispanic	32.8	34.1	-1.4	0.91
Black, non-Hispanic	25.7	34.5	-8.8	0.31
Other race/ethnicity	11.2	7.1	4.2	0.11
Race missing	0.2	0.0	0.2	0.15
Main language spoken at home is English	82.1	86.9	-4.7	0.58
Biological mother living in home or main home	87.7	80.4	7.4	0.07
Biological father living in home or main home	53.5	58.2	-4.7	0.47
Biological parents currently married	49.3	42.4	6.9	0.30
Ever smoked cigarettes	24.4	30.0	-5.7	0.32
Drank alcohol in the past 30 days	26.5	19.0	7.5	0.06
Smoked marijuana in the past 30 days	14.3	11.6	2.6	0.41
Exposure to Teen PEP messages				
Received any information in past 12 months on:				
Relationships	86.9	87.1	-0.2	0.96
Abstinence	68.6	66.9	1.7	0.79
Methods of birth control	49.7	45.9	3.8	0.36
Where to get birth control	44.4	41.3	3.1	0.46
STIs	80.6	80.8	-0.1	0.98
Talking about sex with your partner	51.9	46.8	5.1	0.50
Saying no to sex	75.7	77.5	-1.9	0.64
How babies are made	84.5	85.3	-0.8	0.82
Sexual behaviors				
Ever had sexual intercourse	22.6	20.6	2.0	0.51
Number of lifetime sexual partners	0.73	0.40	0.33*	0.04
In the three months before survey				
Had sexual intercourse ^a	12.7	10.4	2.3	0.31
Had sexual intercourse without a condom ^a	7.8	6.0	1.9	0.31
Had sexual intercourse without any effective method of birth control	6.3	4.4	2.0	0.15
Ever had nonpenetrative sex	49.3	48.4	0.8	0.88
Ever kissed member of opposite sex	83.6	83.3	0.3	0.94
Sample size	783	248		

Source: Teen PEP baseline survey.

Note: Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each intervention school equal weight. *P*-values corrected for clustering at the school level.

^aConfirmatory outcome.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

D. Impacts of Teen PEP on sexual risk behaviors

Across all outcomes and estimation approaches, we find no evidence that Teen PEP reduced sexual risk behaviors within the combined sample of New Jersey and North Carolina schools. All estimated impacts of the program are statistically insignificant and most are substantively small.

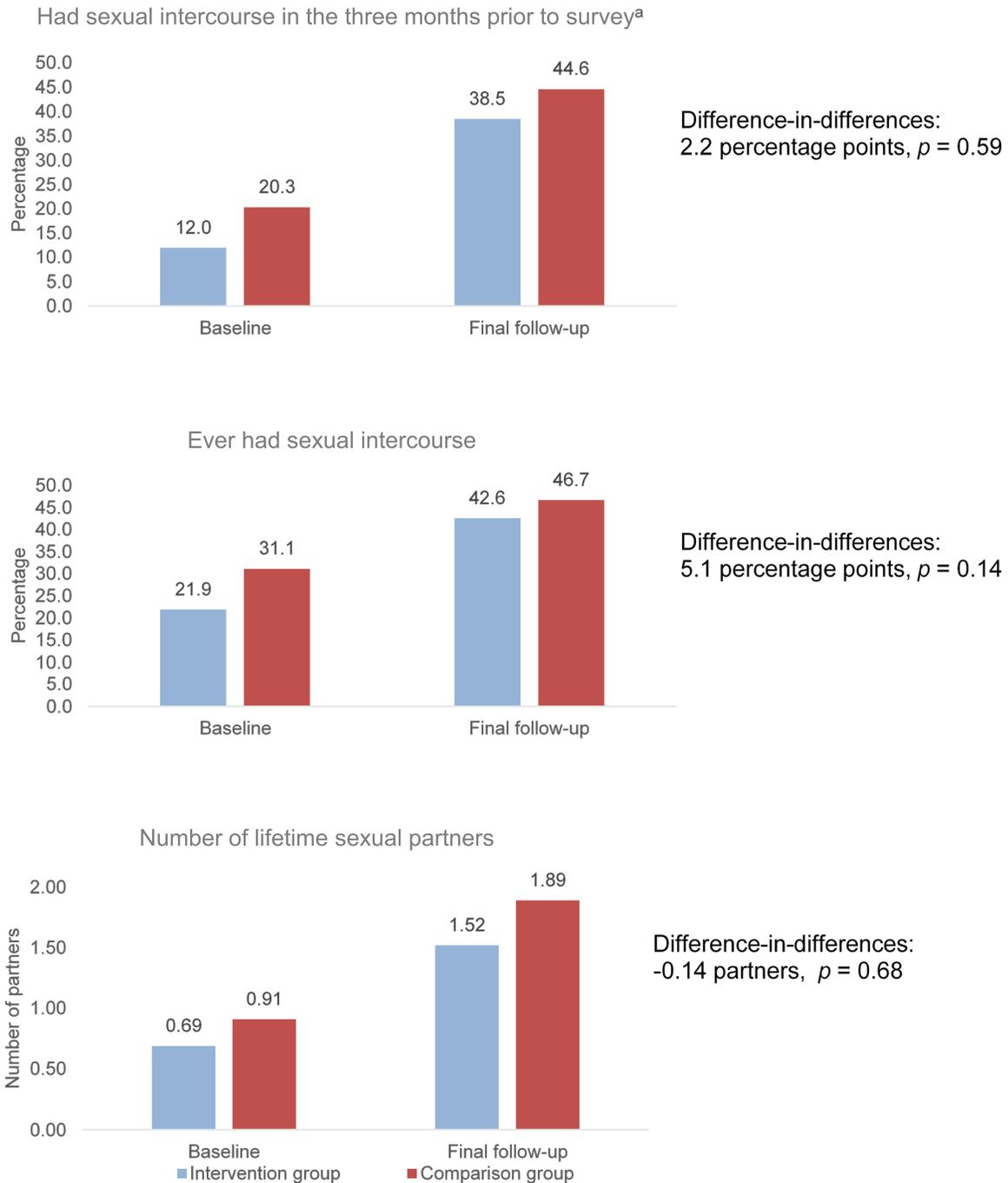
1. Prevalence of sexual activity

For the measures of sexual activity, results from our DD analysis of data for the full sample show that students in intervention schools reported lower rates of sexual activity at both baseline and follow-up, compared to students in comparison schools (Figure 2). For example, at the time of the baseline survey, 12 percent of students attending Teen PEP schools reported having had sexual intercourse in the prior three months, compared to 20 percent of students at comparison schools. Similarly, at the time of the final follow-up survey, 39 percent of students attending Teen PEP schools reported having had sex in the prior three months, compared to 45 percent of students at the comparison schools. Thus, the difference between intervention and comparison schools in rates of sexual activity went from 8 percentage points at baseline to 6 percentage points at follow-up. This implies an insignificant, 2 percentage point DD estimate of the impact of Teen PEP on rates of sex within the past three months ($p = 0.59$).

DD estimates for other outcomes in the sexual activity domain are similarly insignificant (Figure 2). At the time of the baseline survey, 22 percent of students at Teen PEP schools and 31 percent of students at comparison schools reported having ever had sex. At the time of the final follow-up survey, these shares had increased to 43 and 47 percent, respectively, meaning that the difference in rates between the intervention and comparison groups shrank from about 9 percentage points at baseline to 4 percentage points at follow-up. Taking the difference in these differences implies that Teen PEP increased rates of sexual initiation by 5 percentage points; however, the estimate is not statistically significant ($p = 0.14$). The same estimation method also implies that Teen PEP reduced students' reports of lifetime sexual partners by 0.14 partners. However, again, the estimated impact is statistically insignificant ($p = 0.68$).

For the measures of sexual activity, our primary estimation method, and the other statistical methods considered, all confirm this pattern of results (Table 3). Using the propensity-score matched sample for New Jersey and North Carolina schools, which provides the most rigorous estimates of the impact of Teen PEP, implies that the program did not lead to reductions in sexual activity. Within this sample, 38 percent of students reported having had sex in the three months before the final follow-up survey, compared to 41 percent of comparison-group students. The difference is statistically insignificant ($p = 0.60$). We also found no evidence of a statistically significant effect on the reported number of sexual partners (a difference of 0.38 partners, $p = 0.18$) or reported rates of sexual initiation (a difference of less than 1 percentage point in absolute value, $p = 0.94$) within the propensity-score matched sample. Finally, impacts estimated using regression analysis on the full sample or propensity-score trimmed sample were also statistically insignificant.

Figure 2. Impacts of Teen PEP on sexual activity, difference-in-differences in full sample



Source: Teen PEP baseline and final follow-up surveys, full sample.

Notes: The sample includes 890 intervention-group students and 479 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each school equal weight. Estimates are in percentages unless otherwise specified. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

Table 3. Impacts of Teen PEP on sexual activity, estimated by regression analysis

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Had sexual intercourse in the past three months ^a				
Propensity-score matched sample	38.2	40.7	-2.5	0.60
Propensity-score trimmed sample	41.6	40.7	1.0	0.80
Full sample	41.5	40.3	1.1	0.77
Ever had sexual intercourse				
Propensity-score matched sample	45.4	45.7	-0.3	0.94
Propensity-score trimmed sample	49.2	46.5	2.8	0.45
Full sample	49.1	46.0	3.0	0.42
Number of sexual partners				
Propensity-score matched sample	1.49	1.11	0.38	0.18
Propensity-score trimmed sample	1.83	1.49	0.34	0.44
Full sample	1.79	1.50	0.28	0.51

Source: Teen PEP final follow-up survey.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each school equal weight (full and propensity-score trimmed samples) or each intervention school equal weight (propensity-score matched sample). Estimates are in percentages unless otherwise specified. The propensity-score matched sample included 783 intervention-group students and 248 comparison-group students. The propensity-score trimmed sample included 783 intervention-group students and 465 comparison-group students. The full sample included 890 intervention-group students and 479 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

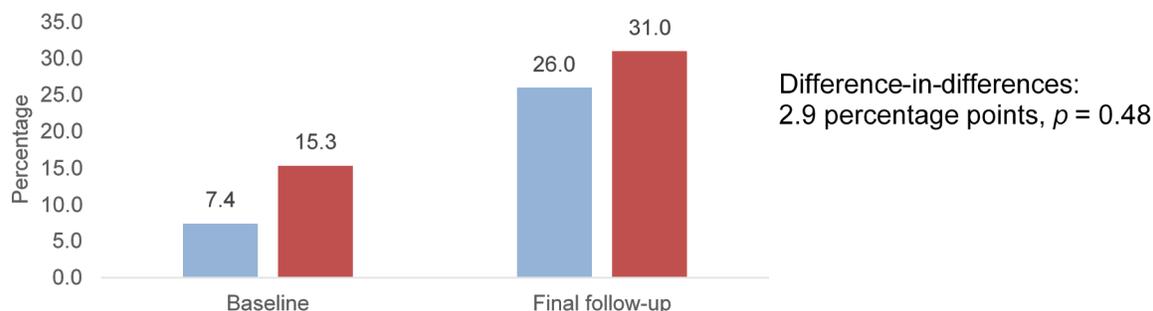
2. Prevalence of unprotected sex

We found similar results for the measures of unprotected sex. For these measures, results from our DD analysis of data for the full sample show students in Teen PEP schools reported lower rates of unprotected sex at both baseline and follow-up, compared to students in the comparison schools (Figure 3). For example, at the time of the baseline survey, 7 percent of students attending Teen PEP schools reported having had sex without a condom in the three months before the baseline survey, compared to 15 percent of students at comparison schools. Similarly, at the time of the final follow-up survey, 26 percent of students attending Teen PEP schools reported having had sex without a condom, compared to 31 percent of students at comparison schools. Thus, the difference in rates between the intervention and comparison groups shrank from 8 percentage points at baseline to 5 percentage points at follow-up. These estimates resulted in a statistically insignificant DD impact estimate of 3 percentage points for this outcome ($p = 0.48$).

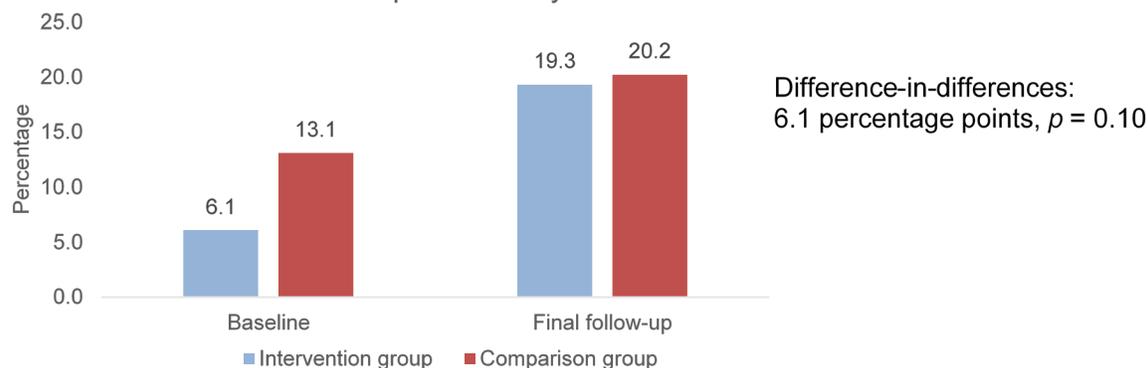
The pattern of estimates is similar for sexual activity without any effective method of birth control. At baseline, 6 percent of Teen PEP students and 13 percent of comparison-group students reported having had sex without any effective birth control method in the past three months. At follow-up, these rates increased to 19 and 20 percent, respectively. The resulting DD impact estimate was 6 percentage points but statistically insignificant ($p = 0.10$).

Figure 3. Impacts of Teen PEP on unprotected sex, difference-in-differences in full sample

Had sexual intercourse without a condom in the three months prior to survey^a



Had sexual intercourse without any effective method of birth control in the three months prior to survey



Source: Teen PEP baseline and final follow-up surveys, full sample.

Notes: The sample includes 890 intervention-group students and 479 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each school equal weight. Estimates are in percentages. P -values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

Impacts produced using more rigorous estimation methods also provide no evidence that Teen PEP influenced the prevalence of unprotected sex in the average school in our sample (Table 4). Within the propensity-score matched sample, 27 percent of both intervention- and comparison-group students reported having engaged in sex without a condom during the three months before the follow-up survey. The small difference between the groups is not statistically significant ($p = 0.89$). Additionally, there is no significant difference within the propensity-score matched sample in the share of students reporting having had sex without any effective method of birth control over the same time period (the difference is less than 1 percentage point in absolute value, $p = 0.96$). Estimates produced using regression analysis on the full sample or propensity-score trimmed sample were similarly small (3 percentage points or less) and insignificant ($p > 0.44$).

Table 4. Impacts of Teen PEP on unprotected sex, estimated by regression analysis

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Had sex without a condom in the past three months ^a				
Propensity-score matched sample	26.6	27.2	-0.7	0.89
Propensity-score trimmed sample	29.2	27.5	1.7	0.63
Full sample	28.3	27.7	0.6	0.87
Had sex without any effective birth control method in past three months				
Propensity-score matched sample	19.6	19.9	-0.3	0.96
Propensity-score trimmed sample	21.2	18.6	2.6	0.44
Full sample	20.4	18.4	2.0	0.55

Source: Teen PEP final follow-up survey.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each school equal weight (full and propensity-score trimmed samples) or each intervention school equal weight (propensity-score matched sample). Estimates are in percentages. The propensity-score matched sample included 783 intervention-group students and 248 comparison-group students. The propensity-score trimmed sample included 783 intervention-group students and 465 comparison-group students. The full sample included 890 intervention-group students and 479 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

E. Discussion and conclusion

This study presents final impact findings from an evaluation of Teen PEP, a comprehensive sexual education program that leverages the power of peer influences to shape adolescent behavior. Teen pregnancy prevention experts have long viewed peer-led programs as a promising approach for reducing teen pregnancy and associated sexual risk behaviors, but research on the effectiveness of these programs has produced mixed results.

Earlier reports on this evaluation suggested that Teen PEP showed promise in changing youth outcomes and behaviors. The analysis of the implementation of Teen PEP revealed that most aspects of the program were well implemented in most study schools and that teens responded positively to program activities (Asheer et al. 2014; Layzer and Rosapep 2012, 2013). Furthermore, our earlier, interim impact analysis revealed that Teen PEP achieved some of its most proximal goals, although it did not affect short-term sexual risk behavior. In particular, approximately six months after the program's conclusion, students in schools that implemented Teen PEP reported greater exposure to a wide range of reproductive health topics, including the major components of the Teen PEP curriculum, and improved knowledge of pregnancy and STI/HIV-transmission prevention (Rotz et al. 2016).

Despite these successes, the final impact findings presented in this report suggest that, within the combined sample of New Jersey and North Carolina schools, Teen PEP did not lead to decreases in the incidence of sexual activity or unprotected sex about 18 months after the program's conclusion. Results for these outcomes were similar 6 months after the program's conclusion (see Rotz et al. 2016). The findings in this report hold when the data are analyzed in

a number of ways (using DD, regression analysis, and propensity-score methods) and across five different measures of student sexual risk behavior.

These results should be viewed with some limitations in mind. A relatively large number of schools dropped out of the study, especially from the New Jersey sample. In addition, among the schools retained in the study, the consent process resulted in slightly less than half of the eligible 9th-grade students participating in the study data collection. As a result, the intervention and comparison groups demonstrated substantial and statistically significant differences on a number of characteristics, including large differences in past sexual experience. To mitigate these differences, we used a range of methodological approaches to analyze the data, including quasi-experimental propensity-score matching methods. These methods helped improve the observed balance between the intervention and comparison groups, but they cannot fully offset the risk of bias introduced by the loss of study schools and relatively low student consent rates.

Additionally, although our propensity-score matching procedure improved the similarity of our intervention and comparison groups, it did so by sacrificing statistical power. This study was designed assuming a sample size of 2,778 students (Smith and Colman 2012), but our analytic sample included only 1,031 students after propensity-score matching. This resulted in much less ability to detect impacts of Teen PEP on all outcomes. Also, the restriction in sample size decreases the generalizability of our estimates. That is, our results are only applicable to the sample of students included in our final analysis and not all students responding to the study surveys or all students in schools participating in the study.

Our findings reflect the specific environments in which Teen PEP was implemented. In particular, estimated impacts are relevant only for students participating in the program during the first year it was implemented at each of the intervention schools. These impacts might change as school officials learn more about Teen PEP, school staff overcome initial implementation challenges and become more skilled in implementing the program, and school stakeholders tailor implementation to the specific nature of their schools. Moreover, our implementation evaluation found that many of the Teen PEP faculty advisors had limited classroom experience, which may have reduced their initial ability to help peer educators learn how to effectively teach the interactive program workshops (Layzer and Rosapep 2012, 2013). A future study of Teen PEP, focusing on schools that have a history of successfully implementing the program, could reveal the importance of these forces.

Additionally, the estimated impacts presented here represent the effects of Teen PEP in the schools in our study and not the effects of the program in the average school implementing Teen PEP. This distinction is important for both states involved in the analysis, though in different ways. In New Jersey, when the study team began recruitment, the program had already been widely available and implemented in many schools. The schools in our analysis had not previously chosen to use the program, which suggests that they could differ systematically in ways that could influence the impact of Teen PEP. For example, they may have already had a successful, comprehensive sexual education program in place. By contrast, our study schools in North Carolina were some of the first in the state to implement Teen PEP. As “early adopters,” these schools naturally experienced some implementation challenges in starting the program, especially given cultural and political differences between North Carolina and New Jersey high

schools. In Appendix D, we further explore whether these and other differences may have led to differences in impacts of Teen PEP across the two states.

Alternatively, our results could suggest that Teen PEP, or peer-driven sexual education programs more broadly, may only be able to influence sexual risk-taking if combined with adult-led activities. In particular, the two peer-driven sexual education programs that have shown the most promise—Sisters, Informing, Healing, Living, Empowering and Safer Choices—use a less extreme approach to harness peer influence than the Teen PEP program model (DiClemente et al. 2004; Coyle et al. 1999, 2001; Basen-Engquist et al. 2001). Within these programs, trained adult facilitators and peers work together to deliver content. This may be a more promising avenue for promoting behavior change than the more strictly peer-led approach of Teen PEP.

Finally, we should note that our findings apply only to the 9th-grade students who received the Teen PEP workshops, not the 11th- and 12th-grade students who served as peer educators. Our focus on the 9th-grade students was driven by the evaluation design and difficulty of identifying a suitable comparison group for the peer educators in comparison schools. However, there are several reasons to expect favorable program effects on these older students. For one, past research suggests peer-led programs might impact peer educators more than other students (Caron et al. 2004; Sieving et al. 2014). In addition, Teen PEP peer educators received the largest “dose” of the program, attending a daily class and sometimes delivering workshops repeatedly to several groups of 9th-grade students. Even if Teen PEP did not impact the sexual behavior of the 9th-grade students receiving the workshops, it might have had large and important impacts on the behavior of the older peer educators.

This page has been left blank for double-sided copying.

REFERENCES

- Asheer, S., E. Kisker, and B. Keating. "Training Teens and Transforming School Culture Through Comprehensive Sex Education: An Implementation Study of Teen PEP." Princeton, NJ: Mathematica Policy Research, 2014.
- Basen-Engquist, K., K. Coyle, G. Parcel, D.B. Kirby, S. Banspach, and S. Carvajal. "Schoolwide Effects of a Multicomponent HIV, STD, and Pregnancy Prevention Program for High School Students." *Health Education and Behavior*, vol. 28, no. 2, 2001, pp. 166–185.
- Caron, F., G. Godin, J. Otis, and L.D. Lambert. "Evaluation of a Theoretically Based AIDS/STD Peer Education Program on Postponing Sexual Intercourse and on Condom Use Among Adolescents Attending High School." *Health Education Research*, vol. 19, no. 2, 2004, pp. 185–197.
- Coyle, K., K. Basen-Engquist, D. Kirby, G. Parcel, S. Banspach, J. Collins, E. Baumler, S. Carvajal, and R. Harrist. "Safer Choices: Reducing Teen Pregnancy, HIV, and STDs." *Public Health Reports*, vol. 116, suppl. 1, 2001, pp. 82–93.
- Coyle, K., K. Basen-Engquist, D. Kirby, G. Parcel, S. Banspach, R. Harrist, E. Baumler, and M. Weil. "Short-Term Impact of Safer Choices: A Multicomponent, School-Based HIV, Other STD, and Pregnancy Prevention Program." *Journal of School Health*, vol. 69, no. 5, 1999, pp. 181–188.
- Crump, R.K., V.J. Hotz, G.W. Imbens, and O.A. Mitnik. "Dealing with Limited Overlap in Estimation of Average Treatment Effects." *Biometrika*, vol. 96, no. 1, 2009, pp. 187–199.
- Dehejia, R.H. and S. Wahba. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American statistical Association*, vol. 94, no. 448, 1999, pp. 1053-1062.
- DiClemente, R., G. Wingood, K. Harrington, D.L. Lang, S.L. Davies, E.W. Hook 3rd, M.K. Oh, R.A. Crosby, V.S Hertzberg, A.B. Gordon, J.W. Hardin, S. Parker, and A. Robillard. "Efficacy of an HIV Prevention Intervention for African American Adolescent Girls: A Randomized Controlled Trial." *Journal of the American Medical Association*, vol. 292, no. 2, 2004, pp. 171–179.
- Funk, M.J., D. Westreich, C. Wiesen, T. Stürmer, M.A. Brookhart, and M. Davidian. "Doubly Robust Estimation of Causal Effects." *American Journal of Epidemiology*, vol. 173, no. 7, 2011, pp. 761–767.
- Hoffman, S.D. "Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy." Washington, DC: The Urban Institute Press, 2008.
- Imbens, G.W. "Matching Methods in Practice: Three Examples." *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 373–419.
- Imbens, G.W. and J.M. Wooldridge. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, vol. 47, no. 1, 2009, pp. 5-86.

- Kann, L., S. Kinchen, S.L. Shanklin, K.H. Flint, J. Kawkins, W.A. Harris, R. Lowry, E.O. Olsen, T. McManus, D. Chyen, L. Whittle, E. Taylor, Z. Demissie, N. Brener, J. Thornton, J. Moore, and S. Zaza. "Youth Risk Behavior Surveillance—United States, 2013." *Morbidity and Mortality Weekly Report Surveillance Summaries*, vol. 63, suppl. 4, 2014, pp. 1–168.
- Layzer, C., and L. Rosapep. "Year 2 Interim Evaluation Report: Teen PEP Cohort 1 Participant Experiences." Cambridge, MA: Abt Associates, 2012.
- Layzer, C., and L. Rosapep. "Year 3 Interim Evaluation Report: Teen PEP Cohort 2 Participant Experiences." Cambridge, MA: Abt Associates, 2013.
- Layzer, C., L. Rosapep, and S. Barr. "A Peer Education Program: Delivering Highly Reliable Sexual Health Promotion Messages in Schools." *Journal of Adolescent Health*, vol. 54, suppl. 3, 2014, pp. S70–S77.
- Martin, J.A., B.E. Hamilton, and S.J. Ventura. "Births: Final Data for 2013." Hyattsville, MD: National Center for Health Statistics, 2015.
- Mitchell, T.D., and M.W. Greene. "School Health Profiles in North Carolina Middle and High Schools." Atlanta, GA: Centers for Disease Control and Prevention, 2011.
- Moulton, B.R. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *The Review of Economics and Statistics*, vol. 72, no. 2, 1990, pp. 334–338.
- National Campaign to Prevent Teen and Unplanned Pregnancy. "Counting It Up: The Public Costs of Teen Childbearing." 2013. Available at <http://thenationalcampaign.org/why-it-matters/public-cost>. Accessed July 23, 2015.
- Perper, K., K. Peterson, and J. Manlove. "Diploma Attainment Among Teen Mothers." Washington, DC: Child Trends, 2010.
- Rotz, D., B. Goesling, M. Crofton, C. Trenholm, J. Manlove, and K. Welti. "Interim Impacts of Teen PEP in New Jersey and North Carolina High Schools." Cambridge, MA: Mathematica Policy Research, 2016.
- Schochet, P.Z., A. Duncan, and J.Q. Easton. "Technical Methods Report: The Estimation of Average Treatment Effects for Clustered RCTs of Education Interventions Technical Methods Report: The Estimation of Average Treatment Effects for Clustered RCTs of Education Interventions." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.
- Sieving, R.E., A.L. McRee, M. Secor-Turner, A.W. Garwick, L.H. Bearinger, K.J. Beckman, B.J. McMorris, and M.D. Resnick. "Prime Time: Long-Term Sexual Health Outcomes of a Clinic-Linked Intervention." *Perspectives on Sexual and Reproductive Health*, vol. 46, no. 2, 2014, pp. 91–100.
- Smith, K., and S. Colman. "Evaluation of Adolescent Pregnancy Prevention Approaches: Design of the Impact Study." Princeton, NJ: Mathematica Policy Research, 2012.

APPENDIX A

DATA AND MEASURES

This page has been left blank for double-sided copying.

This appendix provides detailed information on the survey data collection and measures. We begin by describing the survey design and administration. We then discuss how we constructed the measures of sexual activity and unprotected sex. We end by listing the baseline measures we constructed for this analysis.

A. Survey design and administration

The survey instrument was designed to capture a broad range of measures of family background and demographic characteristics, views and attitudes, sexual activity and other youth risk behaviors, and intentions and aspirations. The Evaluation of Adolescent Pregnancy Prevention Approaches (PPA) research team developed the survey, drawing on items found in well-established surveys such as the National Longitudinal Study of Adolescent Health, National Longitudinal Survey of Youth, Youth Risk Behavior Surveillance System, and National Survey of Family Growth. After compiling all relevant items from these surveys, we identified and prioritized those that best served the objectives of the PPA impact study. In some cases, we had to adapt the questions to fit our primary pencil-and-paper survey mode and the age range of our study sample. Most of these adaptations involved changing wording to make questions easier to understand or simplifying the response categories.

We designed the questionnaire so that only students who reported being sexually experienced were asked sensitive questions related to sexual activity. Specifically, the survey was split into three parts. All students completed Part A, which included only general questions about family background and demographic characteristics, views, attitudes, and knowledge. This part of the survey concluded with a single screening question about sexual experience. For the baseline survey administered to Cohort 1, this question was “Have you ever had sexual intercourse, oral sex, or anal sex?” In response to concerns about the sensitive nature of this question, we changed this screening question to “Have you ever had sexual intercourse?” for the baseline survey of Cohorts 2 and 3 and “Have you ever had sexual intercourse or oral sex?” for the final follow-up survey of all cohorts. In all cases, students who answered “yes” to this screening question were instructed to complete Part B1 of the survey, which contained questions regarding sexual risk behaviors that were more detailed. Students who answered “no” to the screening question were instructed to complete Part B2 of the survey, which included an alternative set of questions. Both Parts B1 and B2 began with a question asking students to confirm their answers to the screening question and instructed the students to either continue completing the form they selected (if the answer was confirmed) or switch to the alternative form (if not). Parts B1 and B2 of the survey were formatted to look indistinguishable, so that when administering the survey in a group setting, students could not tell which part of the survey other respondents were completing.

As is the case with any self-reported survey, the survey responses might be subject to reporting bias. For the final follow-up survey, we were primarily concerned with the questions relating to sexual behavior. For these measures, the reporting bias might occur in either direction. On the one hand, students in the intervention group might be less likely to report risky sexual behaviors because they are embarrassed to admit to a behavior the program discourages. Such underreporting could lead to a spurious finding of *lower* sexual activity among students in the intervention group. On the other hand, especially because our study sample is relatively young, it is possible that the program made students in the intervention group better informed about sexual

risk behaviors and therefore more likely to report their true involvement in these behaviors. Such an effect could lead to a spurious finding of *higher* sexual activity rates among students in the intervention group.

We made several different efforts to minimize these risks. To help encourage honest reporting, independent field staff trained and employed by the study team—not the school teachers or anyone else personally connected to the study participants—administered the survey. At the beginning of each survey administration, these staff reminded students that their answers would be kept confidential and encouraged them to respond truthfully. To help maintain their privacy, students were seated at a comfortable distance from their peers during survey administration and placed completed questionnaires in envelopes before handing them over to the survey staff. Questionnaires and return envelopes were labeled with a unique ID number with no personally identifiable information appearing on either. Questions were asked in an objective manner, and the survey instruments were pretested to ensure that questions were worded appropriately for the study sample.

B. Outcome measures

As discussed in the main body of this report, we used two confirmatory outcomes and three exploratory outcomes to assess whether Teen PEP reduced rates of risky sexual behavior. A binary (yes/no) indicator of whether a student reported having sexual intercourse in the past three months served as our primary measure of whether exposure to Teen PEP leads to a reduction in sexual activity. Other outcomes used to assess this research question were a binary indicator for whether a student reported ever having sexual intercourse and the number of lifetime sexual partners a student reported. A binary indicator of whether a student reported having sexual intercourse without using a condom in the past three months served as the primary measure of whether exposure to Teen PEP leads to a reduction in unprotected sex. We also used a binary indicator of whether a student reported having sexual intercourse without using any effective method of birth control in the past three months to gauge Teen PEP’s impact within this domain.

We followed a two-step process for constructing these outcomes from the responses provided on the study surveys. First, respondents who answered “no” to the sexual experience screening question at the end of Part A of the questionnaire (discussed in Section A of this appendix) were coded as not engaging in any sexual activity. Second, respondents who answered “yes” to the screening question were asked a series of questions that we used to determine the values of each of the aforementioned outcomes. The survey items covered: (1) whether the student ever had intercourse, (2) whether the student had intercourse more than once, (3) how many sexual partners the student had, (4) how many times the student had intercourse in the past three months, (5) how many times the student had intercourse without using a condom in the past three months, and (6) how many times the student had intercourse without using any of the following birth control methods in the past three months: condoms, birth control pills, the shot (Depo-Provera), the patch, the ring (NuvaRing), IUD (Mirena or Paragard), or implants (Implanon).

Students were instructed to skip any irrelevant questions (for example, students who had not had sex multiple times were not asked how many sexual partners they have had), although they often did not do so when filling out pencil-and-paper surveys. In the baseline survey, students

who reported having had sex only one time were not asked to complete questions about recent sexual activity. In these cases, we assumed that the individual did not have sexual intercourse in the three months before the survey unless he or she filled out the survey to indicate otherwise.

In constructing our outcome measures, we accounted for any observed inconsistent or discrepant responses across different items—for example, participants who reported both never having had sex and having had sex in the past three months. To resolve conflicts between the screening question in Section A, the confirmation of this question in Section B1 or B2, and the survey items on sexual activity, we considered students’ responses to all questions taken together. In cases where there was no preponderance of evidence indicating measures should be coded in a particular way, we recoded the outcomes as missing values. In Appendix C, we further explore the robustness of our results to these decisions for handling inconsistent responses.

C. Baseline covariates

We constructed a broad range of measures using data from the baseline survey to assess the equivalence of our study groups, select propensity-score trimmed and matched samples, and adjust for remaining differences between the intervention- and comparison-group students when producing impact estimates. We define these variables in Table A.1.

Table A.1. Baseline characteristics of study sample

Measure	Definition
Demographic and background characteristics	
Age	Continuous variable: approximate age when student completed baseline survey, calculated based on the date when the baseline survey was completed and the student’s reported month and year of birth.
Female	Binary variable: equals 1 if a student is female; equals 0 if a student is male.
Hispanic	Binary variable: equals 1 if a student self-identifies as Hispanic or Latino/a; equals 0 otherwise.
White, non-Hispanic	Binary variable: equals 1 if a student self-identifies as white, non-Hispanic; equals 0 otherwise.
Black, non-Hispanic	Binary variable: equals 1 if a student self-identifies as black, non-Hispanic; equals 0 otherwise.
Other race/ethnicity	Binary variable: equals 1 if a student self-identifies as non-Hispanic and American Indian, Alaskan Native, Asian, Native Hawaiian, or Pacific Islander, or is of multiple race/ethnicities; equals 0 otherwise.
Race missing	Binary variable: equals 1 if a student’s race/ethnicity is missing; equals 0 if a student’s race/ethnicity is nonmissing.
Main language spoken at home is English	Binary variable: equals 1 if English is the main language spoken at the student’s home; equals 0 if English is not the main language spoken at the student’s home.
Biological mother living in home or main home	Binary variable: equals 1 if a student lives with his or her biological mother; equals 0 if a student does not live with his or her biological mother.
Biological father living in home or main home	Binary variable: equals 1 if a student lives with his or her biological father; equals 0 if a student does not live with his or her biological father.

Measure	Definition
Biological parents currently married	Binary variable: equals 1 if a student's biological parents are currently married to each other; equals 0 if a student's biological parents are not currently married to each other.
Biological parents currently divorced or separated	Binary variable: equals 1 if a student's biological parents were once married to each other but are now separated or divorced; equals 0 if a student's biological parents are not currently separated or divorced.
Report religion is very important in their life	Binary variable: equals 1 if a student reported that religion was "very" important in his or her life; equals 0 if a student reported that religion was "somewhat important" or "not at all important."
Identify as lesbian, gay, bisexual, asexual, or questioning	Binary variable: equals 1 if a student reported that he or she is a lesbian, gay, bisexual, asexual, or questioning; equals 0 if a student reported that he or she is heterosexual.
Ever smoked a cigarette	Binary variable: equals 1 if a student reported having ever smoked a cigarette; equals 0 if a student reported never having smoked a cigarette.
Ever drank alcohol	Binary variable: equals 1 if a student reported ever having more than a sip of an alcoholic drink; equals 0 if a student reported never having more than a sip of an alcoholic drink.
Ever smoked marijuana	Binary variable: equals 1 if a student reported having ever smoked marijuana; equals 0 if a student reported never having smoked marijuana.
Measures of sexual behavior	
Ever had sexual intercourse	Binary variable: equals 1 if a student reported that he or she has ever had sexual intercourse; equals 0 if a student reported that he or she has not had sexual intercourse.
Number of lifetime sexual partners	Count variable: equals the total number of sexual partners the participant has ever had. Enters regressions and propensity-score matching as a series of binary variables indicating whether someone had one, two, or three or more lifetime sexual partners.
Had intercourse in the past three months	Binary variable: equals 1 if a student reported that he or she had intercourse at least once in the three months before completing the survey; equals 0 if a student reported that he or she did not have intercourse in the three months before completing the survey.
Had intercourse without a condom in the past three months	Binary variable: equals 1 if a student reported that he or she had intercourse without a condom at least once in the three months before completing the survey; equals 0 if a student reported that he or she did not have intercourse without a condom in the three months before completing the survey.
Had intercourse without using any effective method of birth control in the past three months	Binary variable: equals 1 if a student reported that he or she had intercourse without any effective form of birth control at least once in the three months before completing the survey; equals 0 if a student reported that he or she did not have intercourse without any effective form of birth control in the three months before completing the survey.
Ever had oral sex	Binary variable: equals 1 if a student reported that he or she has ever had oral sex; equals 0 if a student reported that he or she has not had oral sex. This construct is only available at baseline for Cohort 1.
Ever had nonpenetrative sex	Binary variable: equals 1 if a student reported that he or she has ever had nonpenetrative sex; equals 0 if a student reported that he or she has not had nonpenetrative sex.
Ever kissed member of opposite sex	Binary variable: equals 1 if a student reported that he or she has ever kissed a member of the opposite sex; equals 0 if a student reported that he or she has not kissed a member of the opposite sex.

Measure	Definition
Exposure to information	
Received information about relationships	Binary variable: equals 1 if a student reported having received any information about “relationships, dating, marriage, or family life” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Received information about abstinence	Binary variable: equals 1 if a student reported having received any information about “abstinence from sex” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Received information about birth control methods	Binary variable: equals 1 if a student reported having received any information about “methods of birth control” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Received information about where to get birth control	Binary variable: equals 1 if a student reported having received any information about “where to get birth control” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Received information about STIs	Binary variable: equals 1 if a student reported having received any information about “sexually transmitted diseases, also known as STDs” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Received information about talking with your partner about sex	Binary variable: equals 1 if a student reported having received any information about “how to talk to your partner about whether to have sex or whether to use birth control” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Received information about saying no to sex	Binary variable: equals 1 if a student reported having received any information about “how to say no to sex” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Received information about how babies are made	Binary variable: equals 1 if a student reported having received any information about “how babies are made” in the past 12 months; equals 0 if a student reported having not received information on this topic in the past 12 months.
Knowledge	
Knowledge of preventing STI transmission and pregnancy	Continuous index: sum of correct responses to six survey questions; variable ranges from 0 to 6, with higher values indicating greater knowledge. For students who do not respond to all six items, index is set to missing; otherwise, missing responses were counted as incorrect.
Attitudes	
Perceptions of factors supporting use of birth control	Continuous scale: average of responses to four survey questions; variable ranges from 1 to 5, with higher values indicating more supportive attitudes toward birth control.
Perceptions of barriers to use of birth control	Continuous scale: average of responses to five survey questions; variable ranges from 1 to 5, with higher values indicating more perceived barriers toward birth control use.
Negative views toward early sexual activity	Continuous scale: average of responses to four survey questions; variable ranges from 1 to 5, with higher values indicating less permissive attitudes toward early sexual activity.
Would be very upset if got pregnant or got someone pregnant	Binary variable: equals 1 if a student reported he or she would be “very upset” if she became pregnant or he impregnated someone now; equals 0 if a student chose another response category (indicating he or she would be less upset or happy about a pregnancy).
Intentions	
Intend to have sex in the next year	Binary variable: equals 1 if a student reported he or she will “definitely” have sex in the next year if he or she has the chance; equals 0 if a student reported he or she will “probably”, “probably not” or “definitely not” do so.

Measure	Definition
Intend to use a condom if have sex in the next year	Binary variable: equals 1 if a student reported he or she will “definitely” use a condom if he or she has sex in the next year; equals 0 if a student reported he or she will “probably”, “probably not” or “definitely not” do so.
Intend to use any effective method of birth control if have sex in the next year	Binary variable: equals 1 if a student reported he or she will “definitely” use an effective method of birth control if he or she has sex in the next year; equals 0 if a student reported he or she will “probably”, “probably not” or “definitely not” do so.
Decision making and self-efficacy	
Perceptions of refusal skills	Continuous scale: average of two questions for male students and three questions for female students; variable ranges from 1 to 4, with higher values indicating greater perceived refusal skills.
Communication	
Frequency of discussions with parents about relationships or sexual health in past 12 months	Continuous scale: based on seven survey questions; variable ranges from 0 to 10, with higher values indicating more communication.
Spoke to health professional about sexual health in the past 12 months	Binary variable: equals 1 if a student reported he or she spoke with a doctor or nurse about sex, birth control, or sexually transmitted diseases in the past 12 months; equals 0 if a student reported otherwise
Substance use	
Drank alcohol in past 30 days	Binary variable: equals 1 if a student reported drinking alcohol one or more times in the past 30 days; equals 0 if a student reported he or she did not do so.
Smoked marijuana in past 30 days	Binary variable: equals 1 if a student reported smoking marijuana one or more times in the past 30 days; equals 0 if a student reported he or she did not do so.

APPENDIX B

ANALYTIC METHODS

This page has been left blank for double-sided copying.

We designed our analysis methods accounting for two key issues. First, a relatively large number of schools dropped out of the study after random assignment, especially from the New Jersey sample. Because schools left the study in a non-random manner, this school-level attrition has the potential to compromise the validity of the original random assignment evaluation design, and we cannot be confident that any differences between the study groups in our analytic sample at baseline are due only to chance. Second, among the 17 schools that were retained for the study, the baseline characteristics of the student samples differed markedly between the intervention and comparison groups. For example, students in intervention schools were about half as likely as students in comparison schools to report having had sex without using a condom in the three months before the baseline survey. We must adjust for these pre-existing differences when analyzing data from the follow-up survey to avoid making false claims about the program's impacts on student outcomes.

We used four different analytic techniques to produce estimates of the impact of Teen PEP in light of these issues: difference-in-differences (DD), regression analysis on the full sample, regression analysis on a propensity-score trimmed sample, and regression-analysis on a propensity-score matched sample. The DD and full-sample regression analyses use all available data to provide estimates relevant to the broadest possible set of students. The other regression analyses use propensity-score methods to reduce the sample to a smaller set of students, in which the intervention and comparison groups are more similar, providing estimates with greater causal validity.

This appendix describes in detail these four analytic approaches. Section A describes our preferred analytic approach: propensity-score matching combined with regression. Section B details the other methods.

A. Propensity-score matching

Our preferred estimation approach used propensity-score matching to select a subset of intervention- and comparison-group students for our analysis, restricting the study sample to a well-matched subset of the original sample. Selecting a subgroup of individuals for analysis enables us to decrease differences in observed characteristics between the intervention and comparison groups at baseline and therefore improve the causal validity of our impact estimates. However, this decrease in sample means that we have reduced statistical power and can make inferences only for smaller subset of students. That is, our estimates are only valid for the students in our final analysis sample and not the entirety of the students enrolled in the study.

Although we conducted random assignment at the school level, we estimated the propensity score using data at the student level. Because our sample contained only 17 schools, it would have been infeasible to estimate propensity-score regressions using school-level data. Other methods of school-level matching (for example, coarsened exact matching) would have resulted in relatively poor balance between the intervention and comparison groups because of the small number of schools. Because of these constraints, we decided to use propensity-score matching to select a group of students in comparison schools most similar to the students in Teen PEP schools, focusing on the student-level data.

The details of our propensity-score approach follow the recommendations of Imbens (2015). We used a logit model and chose the precise propensity-score regression specification using a

stepwise procedure. First, we pre-specified that age, race, gender, state, cohort, and baseline measures of sexual risk behaviors (indicators for sex and sex without a condom in the past three months) be factors used in matching observations. We then iteratively added variables from most to least predictive of treatment status to the propensity-score regression to determine which other variables should enter the model. We considered a rich set of demographic and background characteristics, baseline measures of outcomes, and related variables in this process (listed in Table B.1). After determining the variables that we would use in the propensity-score model, we also included in the model any first-order interactions of these variables that were highly predictive of treatment status. For all analyses, we imputed any missing covariates to their mean value to avoid limiting sample size.

Table B.1. Candidate and selected variables for propensity-score regression

Variable	Included in propensity-score regression?
Ever had sexual intercourse	
Number of lifetime sexual partners (as indicators for 0,1, 2, and 3 or more)	X
Had sexual intercourse without any effective method of birth control in the three months before survey	X
Ever had oral sex (Cohort 1 only)	X
Ever had nonpenetrative sex	
Ever kissed member of opposite sex	X
Received any information in past 12 months on:	
Relationships	
Abstinence	
Methods of birth control	
Where to get birth control	
STIs	
Talking about sex with your partner	
Saying no to sex	X
How babies are made	X
Knowledge of preventing STI transmission and pregnancy	X
Perceptions of factors supporting use of birth control	
Perceptions of barriers to use of birth control	
Index of negative views toward having sex at current age	X
Would be very upset if got pregnant or got someone pregnant	X
Intend to have sex in the next year	
Intend to use a condom if have sex in the next year	
Intend to use any effective method of birth control if have sex in the next year	
Perceptions of refusal skills	
Spoke to health professional about sexual health in past year	
Communication with parents about risky behavior	
Drank alcohol in past 30 days	X
Smoked marijuana in past 30 days	X
Main language spoken at home is English	X
Biological mother living in home or main home	
Biological father living in home or main home	
Biological parents currently married	X
Biological parents divorced or separated	X
Report religion is very important in their life	X

Variable	Included in propensity-score regression?
Identify as lesbian, gay, bisexual, asexual, or questioning	
Ever drank alcohol	
Ever smoked marijuana	X

Notes: See Appendix A for definitions of all variables. The model also included several interaction terms, which are available upon request.

We estimated the propensity score based on a logit model with predictor variables for age, race, gender, cohort, indicators for sex and sex without a condom in the past three months, and the variables indicated in Table B.1. We removed any students from our sample who had propensity scores less than 0.05 or greater than 0.95, to improve overlap and ensure our sample excluded individuals who are very different from the average individual in either study group (see Crump et al. 2009). After trimming based on the propensity score, the sample contained 783 intervention-group students and 465 comparison-group students.

Among the students in this trimmed sample, we then matched each intervention-group student to the comparison-group student in the same state with the closest propensity score. Comparison-group students could be matched to multiple intervention-group students. The matching comparison-group students became the comparison group we used in our analysis, with observations weighted based on the intervention-group students they matched. Comparison-group students matched to multiple intervention-group students received a weight equal to the sum of the weights of all matched students. We created weights for the intervention group so that each school implementing Teen PEP contributes equal weight to the analysis; the analysis thus yields the effects of Teen PEP in the average intervention school.

After matching, our analytic sample contained 783 intervention-group students and 248 comparison-group students. The large reduction in the size of the comparison group highlights the differences in the samples at baseline; ultimately, we chose only slightly more than half of comparison-group students in the trimmed sample for the propensity-score matched sample.

We used regression analysis on our matched sample to estimate the impact of Teen PEP, controlling for a number of observed outcomes at baseline that might be sources of bias despite the use of propensity-score matching (Funk et al. 2011). For all outcomes, we estimated ordinary least squares regression models, adjusting standard errors to account for clustering at the school level. Our regressions controlled for a baseline measure of the outcomes, an indicator for whether this variable was missing at baseline, our two confirmatory outcome measures (sex and sex without a condom in the past three months), race, gender, age, cohort, and state. We also included controls for variables that were selected by the propensity-score matching procedure (linear terms only) and variables with differences at baseline in the propensity-score matched sample exceeding 0.05 standard deviations: whether the student had sex without any effective method of birth control in the past three months; number of past sexual partners; whether the student ever had oral sex (available for Cohort 1 only) or ever kissed someone of the opposite gender; whether the student lived with his or her biological mother; biological mother's and father's marital status; main language spoken at home; religiosity; information received in the past 12 months on how babies are made, methods of birth control, and saying no to sex; knowledge of STI transmission and pregnancy prevention; attitudes toward having sex at his or

her current age; whether the student would be upset if she became pregnant or he got someone pregnant; communication with parents; whether the student drank alcohol or smoked marijuana in the past month; and whether the student ever smoked marijuana or smoked cigarettes.

We imputed missing covariates to their mean value to avoid limiting sample size. In cases where more than 10 percent of observations had missing data for a variable, we also included in the regression a binary variable equal to one if an individual had an imputed response (and zero otherwise).

B. Alternative analyses

In addition to the propensity-score matching procedure, we estimated impacts using DD, regression analysis on the full sample, and regression analysis on the propensity-score trimmed sample.

To estimate impacts using DD, we compared the difference in an outcome between the intervention and comparison groups at follow-up to the difference in that outcome between the groups at baseline. Mathematically, the DD impact on outcome y can be expressed as

$$(1) \quad \beta_y = (\bar{y}_{Teen\ PEP, follow-up} - \bar{y}_{comparison, follow-up}) - (\bar{y}_{Teen\ PEP, baseline} - \bar{y}_{comparison, baseline}),$$

Where \bar{y} is the mean of outcome y for the specified group and time period. We estimated impacts by DD using the full sample of students who responded to both the baseline and final follow-up surveys. This includes 890 students in intervention schools and 479 students in comparison schools. We weighted observations so that each school in the study received equal weight.

We also estimated the impact of Teen PEP using the full sample of survey respondents and regression analysis. For all outcomes, we estimated ordinary least squares regression models, adjusting standard errors to account for clustering at the school level. The regressions controlled for the same characteristics used in the regressions of our preferred estimation approach (see Section A of this appendix). We again weighted observations so that each school in the study received equal weight.

Additionally, we estimated impacts using regression analysis on the propensity-score trimmed sample. This sample includes 783 intervention-group students and 465 comparison-group students who had estimated propensity scores between 0.05 and 0.95. As in the full sample analyses, the regressions controlled for the same characteristics used in our preferred estimation approach, and we weighted observations so that each school in the study received equal weight.

APPENDIX C

ROBUSTNESS ANALYSIS

This page has been left blank for double-sided copying.

The impact findings presented in this report are derived from a particular set of analytic decisions, ranging from the data-cleaning procedures used to construct the outcome measures to the specification of the regression models. We made these decisions in accordance with established research standards and the particular features of our study design. However, we also investigated the sensitivity of our results to alternative analytic decisions—namely, (1) the specification of the regression and propensity-score models, (2) the methods used to calculate standard errors and statistical significance tests, and (3) the cleaning decisions used to code our outcome variables. In this appendix, we present the results of these robustness analyses.

In almost all cases, our robustness analyses confirm the findings from the main body of the report: in the combined sample of study schools located in North Carolina and New Jersey, Teen PEP is not associated with any statistically significant changes in student sexual activity or unprotected sex. In one of the seven robustness analyses presented in this appendix, we observe a marginally significant difference in one of our two confirmatory outcomes and significant differences in several exploratory outcomes. These estimates show higher rates of sexual activity and unprotected sex among students in the intervention group; however, the preponderance of evidence suggests that the program had a null impact on the overall study sample.

A. Alternative model specifications

As discussed in the main body of the report, our preferred estimation approach limited the sample of students to a well-matched subset of the original sample using propensity-score matching and estimated impacts on the reduced sample using regression analysis. We also explored the robustness of our findings to three alternative estimation procedures: using differences-in-differences, regression analysis on the full sample of survey respondents, and regression analysis on a sample of survey respondents trimmed based on the propensity score.

To explore the sensitivity of our results to the specific methods we used for the propensity-score analysis, we first considered using a more parsimonious set of regression covariates. Our regressions attempted to control for a large number of variables. Although this can lead to increased precision and decreased bias, there might be a concern that we are overfitting the regression model. Table C.1 thus contains the impact estimates obtained from using fewer covariates in the regression, in particular, controlling only for a baseline measure of the outcome, an indicator for whether this variable is missing, and the linear variables used in the propensity-score matching procedure. Overall, results are similar to our main findings: there is little evidence that, on average, Teen PEP led to a decrease in sexual activity or unprotected sex. Students in intervention schools were 3 percentage points less likely than students in comparison schools to report having had sex in the three months before the survey and 1 percentage point less likely to report having had sex without a condom during that period, but neither estimate is statistically significant ($p > 0.55$).

Table C.1. Impacts of Teen PEP on outcomes, regression only controls for linear terms from propensity score regression

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Sexual activity				
Had sexual intercourse in the three months before survey ^a	38.1	40.8	-2.7	0.55
Ever had sexual intercourse	45.6	45.5	0.1	0.99
Number of sexual partners	1.52	1.09	0.43	0.13
Unprotected sex				
Had sexual intercourse without a condom in the three months before survey ^a	26.2	27.6	-1.3	0.78
Had sexual intercourse without any effective method of birth control in the three months before survey	19.4	20.1	-0.7	0.90
Sample size	783	248		

Source: Teen PEP follow-up survey, propensity-score matched sample.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics, including a baseline measure of the outcome; an indicator for whether this variable is missing; cohort; state; whether the student had sex, had sex without a condom, or had sex without any method of birth control in the past three months; number of past sexual partners; whether the student ever had oral sex or kissed someone of the opposite sex; main language spoken at home; gender; race; age; biological mother's and father's marital status; religiosity; information received in the past 12 months on how babies are made and saying no to sex; knowledge of STI/HIV transmission and pregnancy prevention; attitudes toward sex; whether the student would be upset if he or she became pregnant or got someone pregnant; communication with parents; whether the student ever smoked marijuana or did so in the past month; whether the student drank alcohol in the past month; and indicators for students not reporting information on whether they ever engaged in nonpenetrative sex and whether the student ever drank alcohol or smoked marijuana or did so in the past month. Estimates are weighted to give each intervention school equal weight. Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

We also explored using different functional forms to estimate the impacts of Teen PEP. For our main regression analysis with our propensity-score matched sample, we used ordinary least squares when analyzing all outcomes. In this robustness analysis, we instead used a logit model to analyze binary outcomes. We also replaced the one non-binary outcome we consider, number of sexual partners, with an alternative binary outcome equal to one if a student has had more than one sexual partner and zero otherwise. Thus, this analysis uses logit models to examine the impacts of Teen PEP for all outcomes it considers.

The results are very similar to those seen for the main analysis (Table C.2). Within the sample of all study schools, Teen PEP is not associated with any significant changes in sexual activity or unprotected sex. Impacts on all confirmatory and exploratory variables are less than 3 percentage points and statistically insignificant ($p > 0.52$).

Table C.2. Impacts of Teen PEP on outcomes, logit model used for all outcomes

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Sexual activity				
Had sexual intercourse in the three months before survey ^a	38.0	40.8	-2.8	0.52
Ever had sexual intercourse	52.2	52.4	-0.2	0.96
Have had multiple sexual partners	22.9	22.2	0.7	0.83
Unprotected sex				
Had sexual intercourse without a condom in the three months before survey ^a	26.8	27.0	-0.3	0.95
Had sexual intercourse without any effective method of birth control in the three months before survey	19.8	19.7	0.0	1.00
Sample size	783	248		

Source: Teen PEP first follow-up survey, propensity-score matched sample.

Note: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each intervention school equal weight. Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

Our main analytic approach used a single propensity-score regression to estimate the propensity score. We also explored estimating the propensity-score regression separately for students in North Carolina and New Jersey, as different factors might be important in predicting treatment status across states.

The results of this analysis show higher overall rates of risky sexual behaviors among students in Teen PEP schools (Table C.3). Under this analysis, students in Teen PEP schools were, on average, 5 percentage points more likely to have had sexual intercourse and 6 percentage points more likely to have had sex without a condom in the three months before the follow-up survey, compared to comparison-group students. Neither estimate is significant at the 5 percent level, but the impact on sex without a condom is marginally statistically significant ($p = 0.05$). Moreover, in Teen PEP schools in both North Carolina and New Jersey, 40 percent of students reported ever having had sex, compared to 32 percent of students in comparison schools. The difference of 8 percentage points is statistically significant ($p = 0.04$). Likewise, students in intervention schools reported 1.47 lifetime sexual partners on average, compared to 0.93 lifetime sexual partners reported by students in comparison schools (a difference of 0.53, $p = 0.04$). Finally, students in the intervention group were, on average, 7 percentage points more likely to have reported having had sex without any effective birth control method in the past three months ($p < 0.01$), with 17 percent of intervention-group students and 10 percent of comparison-group students reporting that they engaged in this activity.

Table C.3. Impacts of Teen PEP on outcomes, estimate propensity score models by state

Variable	Intervention group mean	Comparison group mean	Difference	<i>p</i> -value
Sexual activity				
Had sexual intercourse in the three months before survey ^a	35.9	31.1	4.8	0.24
Ever had sexual intercourse	39.6	31.9	7.7*	0.04
Number of sexual partners	1.47	0.93	0.53*	0.04
Unprotected sex				
Had sexual intercourse without a condom in the three months before survey ^a	23.5	18.0	5.5	0.05
Had sexual intercourse without any effective method of birth control in the three months before survey	17.2	10.1	7.1**	<0.01
Sample size	653	204		

Source: Teen PEP follow-up survey, sample matched using propensity score estimated by state.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each intervention school equal weight. Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level.

^aConfirmatory outcome.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

B. Alternative estimates of standard errors

For all analyses presented in the main body of this report, we adjusted the statistical significance tests (*p*-values) to account for the study's cluster random assignment design. That is, because schools (and not students) were randomly assigned to the intervention and comparison groups, regression error terms might be correlated across students in the same school (Schochet et al. 2009). More generally, there is always concern about correlations in regression error terms when one variable (in this case, treatment status) does not vary across individuals in a specific group (in this case, those in the same school; see Moulton 1990).

To examine the sensitivity of our results to this adjustment, we estimated the impact of Teen PEP using our propensity-score matched sample but did not adjust *p*-values for clustering (Table C.4). The findings from this analysis largely mirror those from our preferred estimation approach. Impacts of Teen PEP on all outcomes remain statistically insignificant at the 5 percent level. The *p*-value associated with the difference across study groups in average number of sexual partners changes the most when we do not adjust for clustering. In our main analysis, this *p*-value was 0.18; without clustering, it falls to 0.06, suggesting a marginally significant effect.

Table C.4. Impacts of Teen PEP on outcomes, not adjusting for clustering

Variable	Intervention group mean	Comparison group mean	Difference	<i>p</i> -value
Sexual activity				
Had sexual intercourse in the three months before survey ^a	38.2	40.7	-2.5	0.55
Ever had sexual intercourse	45.4	45.7	-0.3	0.93
Number of sexual partners	1.49	1.11	0.38	0.06
Unprotected sex				
Had sexual intercourse without a condom in the three months before survey ^a	26.6	27.2	-0.7	0.85
Had sexual intercourse without any effective method of birth control in the three months before survey	19.6	19.9	-0.3	0.94
Sample size	783	248		

Source: Teen PEP first follow-up survey, propensity-score matched sample.

Note: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each intervention school equal weight. Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

C. Alternative data-cleaning procedures

As described in Appendix A, our analysis of the self-reported survey data uncovered some inconsistent or discrepant responses to the questions on sexual risk behaviors. For example, a participant might have reported not having had sex in the past three months but having had sex without a condom over the same period. For the main impact findings presented in this report, we accounted for these discrepancies when creating our outcome measures by considering the preponderance of evidence across all relevant questions in the survey. However, we also examined the sensitivity of our results to three alternative methods for cleaning the data:

- **Alternative coding 1:** Coding a participant as having engaged in a specific behavior if *any* survey item indicates he or she did so.
- **Alternative coding 2:** Coding a participant as *not* having engaged in a specific behavior if any survey item indicates he or she did *not* do so.
- **Alternative coding 3:** Dropping a participant from the analysis if the survey items show a pattern of inconsistent responses.

The results of these analyses show that our findings for the combined sample of New Jersey and North Carolina schools are robust to the specific data-cleaning procedures we used to construct the outcomes (Table C.5–C.7). Using the propensity-score matched sample across all outcomes and all alternative codings, the differences in the behaviors of students at the Teen PEP and comparison schools in our study are statistically insignificant ($p > 0.25$).

Table C.5. Impacts of Teen PEP on outcomes, alternative coding of sexual behaviors 1

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Sexual activity				
Had sexual intercourse in the three months before survey ^a	38.4	40.7	-2.3	0.63
Ever had sexual intercourse	50.2	50.2	0.0	0.99
Number of sexual partners	1.43	1.11	0.33	0.25
Unprotected sex				
Had sexual intercourse without a condom in the three months before survey ^a	26.7	29.2	-2.5	0.60
Had sexual intercourse without any effective method of birth control in the three months before survey	19.6	20.5	-0.9	0.86
Sample size	783	248		

Source: Teen PEP first follow-up survey, propensity-score matched sample.

Note: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each intervention school equal weight. Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

Table C.6. Impacts of Teen PEP on outcomes, alternative coding of sexual behaviors 2

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Sexual activity				
Had sexual intercourse in the three months before survey ^a	36.0	40.9	-4.9	0.26
Ever had sexual intercourse	40.6	41.0	-0.4	0.93
Number of sexual partners	1.40	1.14	0.26	0.29
Unprotected sex				
Had sexual intercourse without a condom in the three months before survey ^a	24.7	26.7	-2.0	0.67
Had sexual intercourse without any effective method of birth control in the three months before survey	18.8	19.8	-0.9	0.85
Sample size	783	248		

Source: Teen PEP first follow-up survey, propensity-score matched sample.

Note: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each intervention school equal weight. Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

Table C.7. Impacts of Teen PEP on outcomes, alternative coding of sexual behaviors 3

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Sexual activity				
Had sexual intercourse in the three months before survey ^a	37.1	40.6	-3.4	0.46
Ever had sexual intercourse	45.0	44.5	0.4	0.91
Number of sexual partners	1.44	1.11	0.33	0.24
Unprotected sex				
Had sexual intercourse without a condom in the three months before survey ^a	25.4	27.2	-1.8	0.71
Had sexual intercourse without any effective method of birth control in the three months before survey	19.0	19.9	-0.9	0.86
Sample size	783	248		

Source: Teen PEP first follow-up survey, propensity-score matched sample.

Note: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each intervention school equal weight. Estimates are in percentages unless otherwise specified. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

This page has been left blank for double-sided copying.

APPENDIX D

STATE-SPECIFIC ESTIMATES

This page has been left blank for double-sided copying.

In this appendix, we explore whether the impacts of Teen PEP vary by state, estimating effects separately for New Jersey and North Carolina schools. The analysis is valuable for both analytic and contextual reasons. School-level attrition was much lower in North Carolina than in New Jersey. Of the 17 New Jersey schools that the study team randomly assigned, 5 remained in the study, compared to 12 of 13 randomly assigned North Carolina schools. This suggests that an analysis of North Carolina schools alone may have greater causal validity than an analysis of North Carolina and New Jersey schools together. Additionally, the context of program implementation differed across the two states. Teen PEP was designed in New Jersey with the needs of New Jersey schools in mind. The program may need further adaptation to be successful in North Carolina to the extent that the cultural and political climate of schools differed across states. Furthermore, the New Jersey schools involved in the evaluation chose to adopt Teen PEP but had not previously done so, despite the program's relatively wide use across that state. In contrast, the North Carolina schools in the evaluation were "early adopters" of the program within their state. This suggests the schools in the two states may have had different needs or desires for sexual education programs.

Additionally, services and programs available to students in comparison schools differed by state (Asheer et al. 2014; Layzer and Rosapep 2012, 2013), which could also lead to differences in estimated impacts. Students in most New Jersey schools take a health class that includes comprehensive sexual education. In contrast, few comparison schools in the North Carolina sample had comprehensive sexual education during their health class. Moreover, access to reproductive health services varied greatly across schools, with students in North Carolina schools typically having more limited access than students in New Jersey.

Despite the benefits of a state-specific analysis, one should view the results produced by this analysis with caution. This study was not designed with these subgroup analyses in mind; therefore, this analysis should be considered exploratory. Additionally, the relatively small sample size available for each subgroup analysis means we have limited statistical power to detect impacts within these smaller sets of students. This is especially true for the New Jersey analysis. After propensity-score matching, the sample of students includes only 55 students in the New Jersey comparison group and 217 students in the New Jersey intervention group.

Overall, the state-specific analyses suggest Teen PEP may be more successful at achieving its goals in New Jersey than in North Carolina. In North Carolina, the evidence suggests that Teen PEP led to increases in sexual risk behavior, particularly unprotected sex. Conversely, in New Jersey, the evidence suggests that Teen PEP led to decreases in sexual risk behavior. In both states, results were less robust than those observed for the main impact analysis, suggesting more evidence is needed before we draw definitive conclusions about the impacts of Teen PEP in either New Jersey or North Carolina high schools alone.

A Impacts of Teen PEP in North Carolina

The results for North Carolina suggest some evidence of adverse program impacts among students in that state. Our primary approach using propensity-score matching and regression suggests the program increased the average number of sexual partners and rates of sex without a condom and sex without any effective method of birth control. The estimated impact on sex

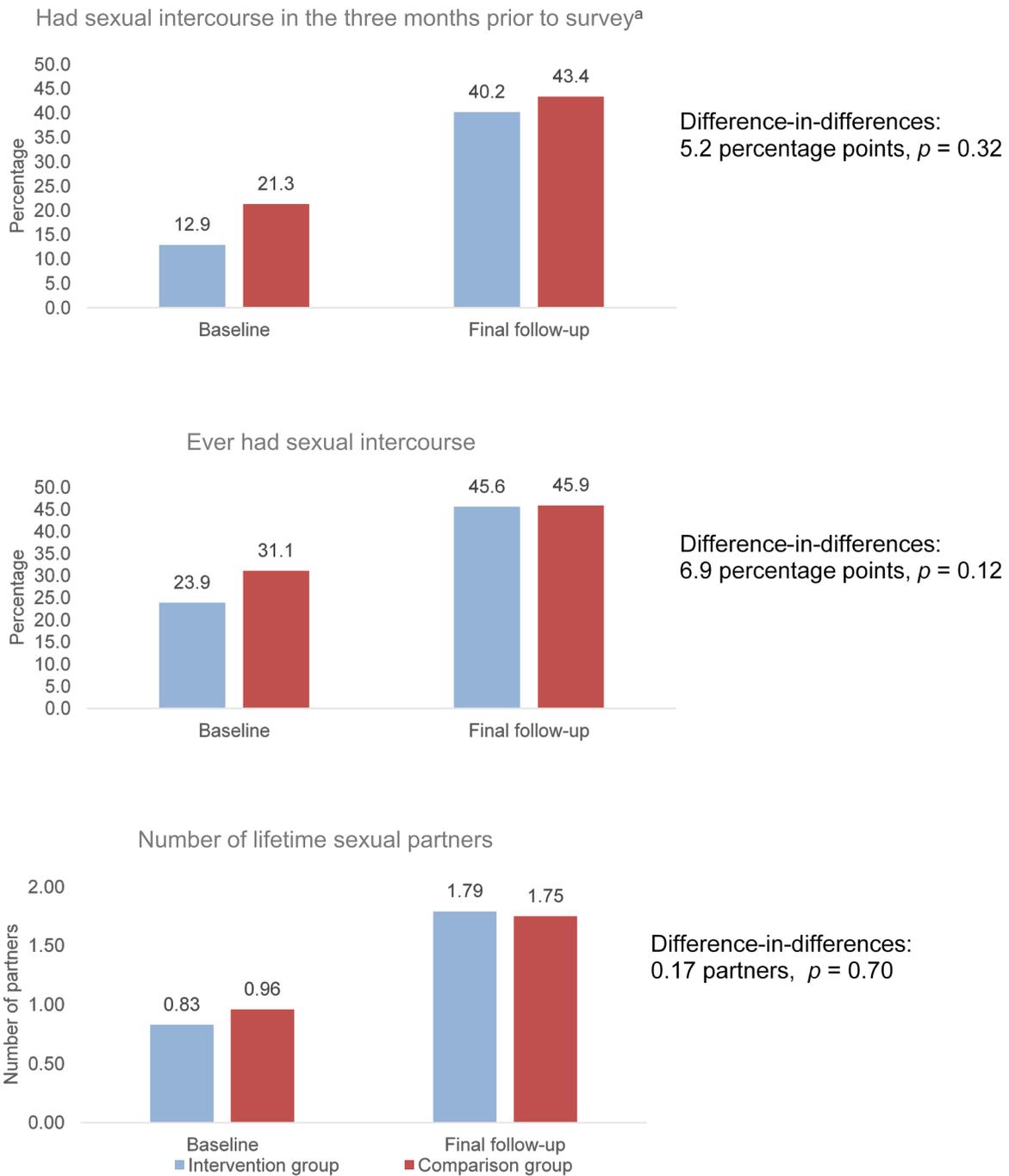
without any effective method of birth control is robust to the alternative specifications considered but impacts estimated for other outcomes are not.

Within the full North Carolina sample, rates of sexual activity were lower in Teen PEP schools than in comparison schools at baseline but more similar at the time of the final follow-up survey (Figure D.1). At baseline, 13 percent of students in North Carolina intervention schools reported having had sexual intercourse in the past three months, compared to 21 percent of students in North Carolina comparison schools. At follow-up, these rates had increased to 40 and 43 percent, respectively. As a result of the changes, we estimated that Teen PEP had an insignificant, 5 percentage point impact on rates of recent sexual activity using differences-in-differences (DD) ($p = 0.32$). Similarly, the share of students who reported ever having sex and students' numbers of reported sexual partners started out lower in North Carolina intervention schools than in North Carolina comparison schools. Both measures increased across the intervention and comparison groups, though they grew by more in intervention schools. As a result, the impacts estimated using DD are positive but statistically insignificant.

When we use our most rigorous analytic method, propensity-score matching and regression, the results mirror those estimated using DD for two of the three measures in the sexual activity domain (Table D.1). About 18 months after the Teen PEP program concluded, students in North Carolina intervention schools were more likely to have had sexual intercourse in the past three months and to have ever had sex than were students in North Carolina comparison schools. But the differences were not statistically significant at the 5 percent level ($p > 0.08$). Differently, impacts on the number of sexual partners become large and significant when we use the propensity-score matching approach. The average student in the North Carolina intervention schools reported 1.78 lifetime sexual partners at the time of our 18-month follow-up survey, compared to 0.98 partners reported by the average North Carolina comparison school student. The 0.80 partner difference is statistically significant ($p = 0.02$).

Although the result for number of sexual partners is suggestive, this estimated impact is not robust to using other analytic methods (Table D.1). Regression analyses using the full or propensity-score trimmed samples imply that the impact on number of sexual partners is statistically insignificant ($p > 0.13$). Moreover, these alternative methods also suggest Teen PEP did not impact the share of students who had sexual intercourse in the three months before the follow-up survey or the share of students who ever had sex. Therefore, the evidence that Teen PEP impacted sexual activity is relatively limited.

Figure D.1. Impacts of Teen PEP on sexual activity, difference-in-differences in full North Carolina sample



Source: Teen PEP baseline and final follow-up surveys, full sample.

Notes: The sample includes 648 intervention-group students and 383 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each school equal weight. Estimates are in percentages unless otherwise specified. P -values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

Table D.1. Impacts of Teen PEP on sexual activity estimated by regression analysis, North Carolina sample

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Had sexual intercourse in the past three months ^a				
Propensity-score matched sample	40.3	35.8	4.5	0.31
Propensity-score trimmed sample	43.4	39.0	4.4	0.39
Full sample	43.4	39.3	4.1	0.41
Ever had sexual intercourse				
Propensity-score matched sample	48.4	42.0	6.3	0.08
Propensity-score trimmed sample	51.9	45.1	6.8	0.16
Full sample	51.7	45.6	6.1	0.20
Number of sexual partners				
Propensity-score matched sample	1.78	0.98	0.80	0.02*
Propensity-score trimmed sample	2.09	1.32	0.77	0.18
Full sample	2.15	1.28	0.86	0.13

Source: Teen PEP final follow-up survey.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each school equal weight (full and propensity-score trimmed samples) or each intervention school equal weight (propensity-score matched sample). Estimates are in percentages unless otherwise specified. The propensity-score matched sample included 566 intervention-group students and 193 comparison-group students. The propensity-score trimmed sample included 566 intervention-group students and 370 comparison-group students. The full sample included 648 intervention-group students and 383 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level.

^aConfirmatory outcome.

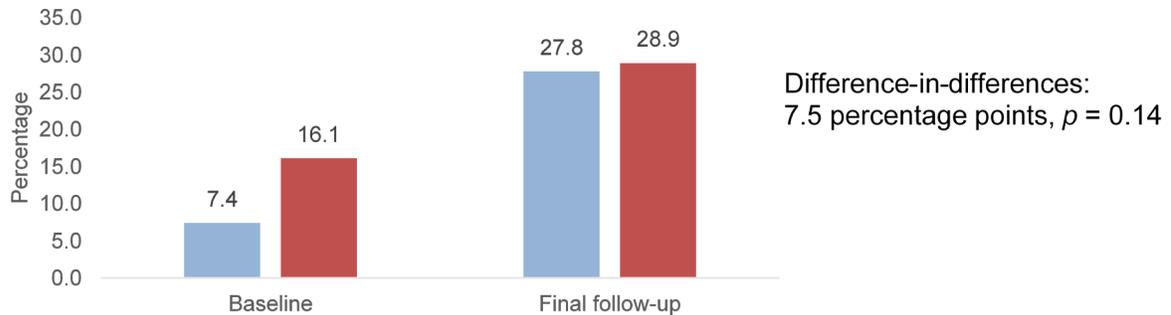
*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

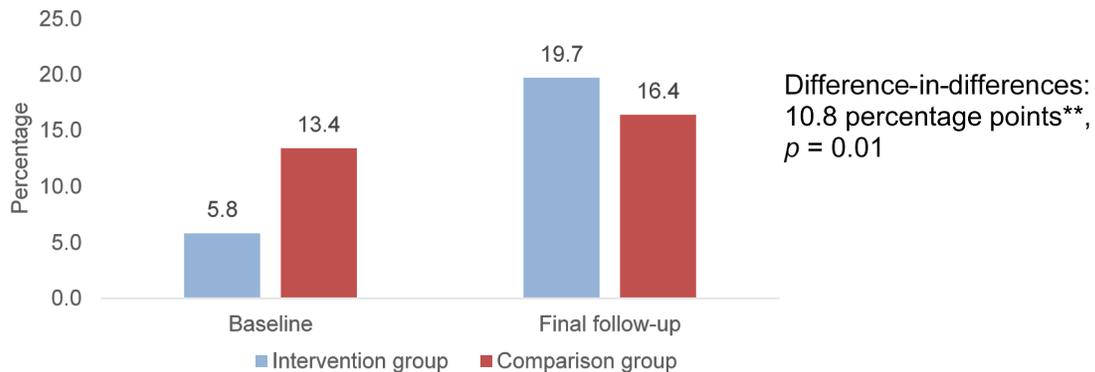
There is more evidence for adverse impacts on North Carolina students in the unprotected sex domain. The pattern of results can be seen from the DD analysis (Figure D.2). At the time of the baseline survey, students in intervention schools were less likely to have had sex without a condom in the past three months than students in comparison schools (7 versus 16 percent). But 18 months after the conclusion of the Teen PEP program, rates of sex without a condom were much more similar in intervention and comparison schools; 28 percent of students in the North Carolina schools implementing Teen PEP reported having had sex without a condom in the past three months at this point, as did 29 percent of comparison-group students. The resulting DD impact estimate is 8 percentage points but statistically insignificant ($p = 0.14$). The results for sex without any effective method of birth control within the past three months were stronger. Compared to the state-specific comparison group, students in North Carolina Teen PEP schools reported lower rates of recent sex without any effective method of birth control at baseline (6 versus 13 percent) but higher rates at the time of the follow-up survey (20 versus 16 percent). Using DD thus implies that Teen PEP was associated with an 11 percentage point increase in rates of sex without any effective method of birth control within the North Carolina sample, a statistically significant impact ($p = 0.01$).

Figure D.2. Impacts of Teen PEP on unprotected sex, difference-in-differences in full North Carolina sample

Had sexual intercourse without a condom in the three months prior to survey^a



Had sexual intercourse without any effective method of birth control in the three months prior to survey



Source: Teen PEP baseline and final follow-up surveys, full sample.

Notes: The sample includes 890 intervention-group students and 479 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each school equal weight. Estimates are in percentages. P -values are corrected for clustering at the school level.

^aConfirmatory outcome.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

The results implied by our propensity-score matching approach are similar (Table D.2). The estimates produced by this analysis imply that students in North Carolina Teen PEP schools were 10 percentage points more likely than students in North Carolina comparison schools to report both having had sex without a condom and having had sex without any effective method of birth control in the past three months. The differences for both outcomes are statistically significant ($p < 0.01$).

Table D.2. Impacts of Teen PEP on unprotected sex estimated by regression analysis, North Carolina sample

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Had sex without a condom in the past three months ^a				
Propensity-score matched sample	28.2	18.7	9.5	0.01**
Propensity-score trimmed sample	30.5	25.0	5.5	0.17
Full sample	31.6	24.7	6.8	0.08
Had sex without any effective birth control method in past three months				
Propensity-score matched sample	20.6	10.3	10.2	<0.01**
Propensity-score trimmed sample	21.6	13.7	8.0	0.04*
Full sample	22.4	13.7	8.7	0.02*

Source: Teen PEP final follow-up survey.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each school equal weight (full and propensity-score trimmed samples) or each intervention school equal weight (propensity-score matched sample). Estimates are in percentages. The propensity-score matched sample included 566 intervention-group students and 193 comparison-group students. The propensity-score trimmed sample included 566 intervention-group students and 370 comparison-group students. The full sample included 648 intervention-group students and 383 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level.

^aConfirmatory outcome.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

When other analytic methods are used (regression on the full sample or regression on the propensity-score trimmed sample, Table D.2), impacts on outcomes in the unprotected sex domain decrease slightly. Using these methods, we estimated that students in North Carolina intervention schools were 6 to 7 percentage points more likely than comparison-group students in the same state to have had sex without a condom in the three months before the 18-month follow-up survey. The differences were not statistically significant at the 5 percent level ($p > 0.08$). The estimated differences in rates of sex without any effective method were somewhat larger (8 to 9 percentage points) and statistically significant ($p < 0.04$).

B. Impacts of Teen PEP in New Jersey

In contrast to the results for the North Carolina sample, our analysis suggests that Teen PEP reduced sexual risk behaviors for New Jersey students. Our most rigorous analytic approach provides evidence that Teen PEP reduced both sexual activity and unprotected sex in New Jersey schools, producing estimated impacts that are both large in magnitude and statistically significant. However, other estimation methods suggest somewhat weaker results in the sexual activity domain. Moreover, the small sample of New Jersey students used in this analysis suggests much care should be taken in interpreting the results.

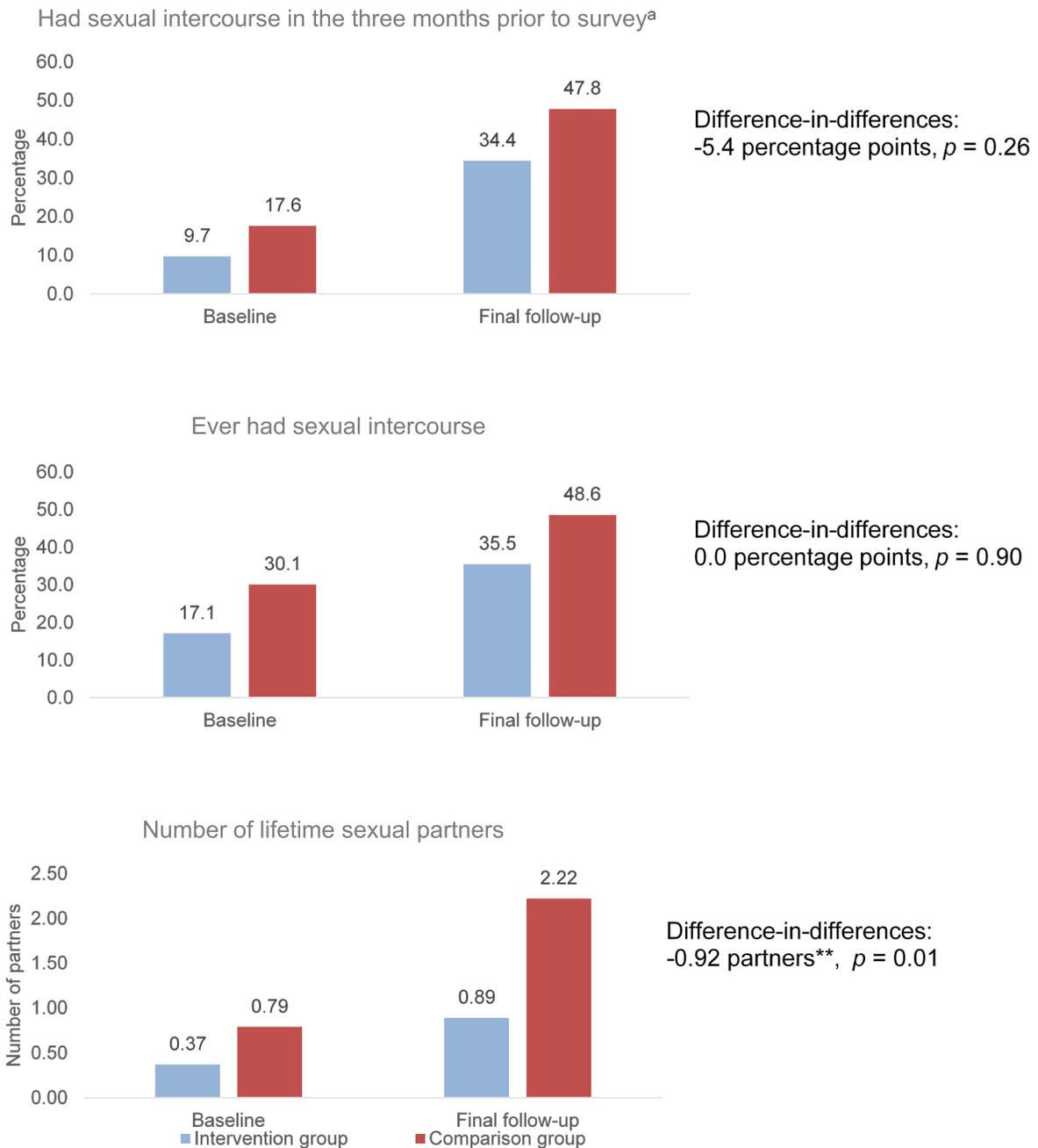
Rates of sexual activity started lower in New Jersey Teen PEP schools than in comparison schools and remained lower throughout the duration of the study (Figure D.3). At baseline, 10 percent of students in intervention schools and 18 percent of students in comparison schools reported having had sex in the past three months. This 8 percentage point gap grew to 13 percentage points by the time of the 18-month follow-up survey. The resulting DD impact

estimate was -5 percentage points but statistically insignificant ($p = 0.26$). The impact on sexual initiation estimated using DD was also small (less than 1 percentage point) and statistically insignificant ($p = 0.90$) within the New Jersey sample. Differently, the gap in the average number of sexual partners reported by students at New Jersey Teen PEP and comparison schools grew over time. At baseline, students in Teen PEP schools reported an average of 0.37 lifetime sexual partners. This increased to 0.89 partners 18 months after the program's conclusion. In comparison schools, the average number of lifetime sexual partners increased from 0.79 to 2.22 over the same period. The resulting DD impact is large (-0.92 partners) and statistically significant ($p = 0.01$).

The results estimated using our most rigorous estimation approach further suggest that Teen PEP led to reductions in sexual activity within New Jersey schools (Table D.3). Focusing on the confirmatory outcome, 34 percent of students at New Jersey Teen PEP schools reported having had sex in the three months before the follow-up survey, compared to 52 percent of students in comparison schools. The 18 percentage point difference is statistically significant ($p < 0.01$). Likewise, this estimation method suggests Teen PEP reduced sexual initiation by 16 percentage points ($p = 0.01$) and the average number of sexual partners by about half ($p = 0.03$) within New Jersey schools.

Although these effects are large and significant, they are not robust to the sample we use to estimate impacts (Table D.4). Regression analyses on the full or propensity-score trimmed samples both imply that students in New Jersey Teen PEP schools were 10 percentage points less likely than students in comparison schools to have had sexual intercourse in the past three months; however, the estimates are not statistically significant ($p > 0.09$), likely because of the small sample of New Jersey students. Impacts on sexual initiation and number of sexual partners estimated using these methods were also sizable but insignificant at the 5 percent level ($p > 0.09$).

Figure D.3. Impacts of Teen PEP on sexual activity, difference-in-differences in full New Jersey sample



Source: Teen PEP baseline and final follow-up surveys, full sample.

Notes: The sample includes 242 intervention-group students and 96 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each school equal weight. Estimates are in percentages unless otherwise specified. P -values are corrected for clustering at the school level.

^aConfirmatory outcome.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

Table D.3. Impacts of Teen PEP on sexual activity estimated by regression analysis, New Jersey sample

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Had sexual intercourse in the past three months ^a				
Propensity-score matched sample	34.0	51.7	-17.7	<0.01**
Propensity-score trimmed sample	36.3	45.8	-9.5	0.14
Full sample	35.6	45.9	-10.3	0.09
Ever had sexual intercourse				
Propensity-score matched sample	38.4	54.3	-15.9	0.01**
Propensity-score trimmed sample	42.6	49.5	-6.9	0.21
Full sample	41.6	49.4	-7.8	0.13
Number of sexual partners				
Propensity-score matched sample	0.82	1.44	-0.61	0.03*
Propensity-score trimmed sample	1.09	1.99	-0.89	0.10
Full sample	1.06	1.96	-0.90	0.09

Source: Teen PEP final follow-up survey.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each school equal weight (full and propensity-score trimmed samples) or each intervention school equal weight (propensity-score matched sample). Estimates are in percentages unless otherwise specified. The propensity-score matched sample included 217 intervention-group students and 55 comparison-group students. The propensity-score trimmed sample included 217 intervention-group students and 95 comparison-group students. The full sample included 242 intervention-group students and 96 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level.

^aConfirmatory outcome.

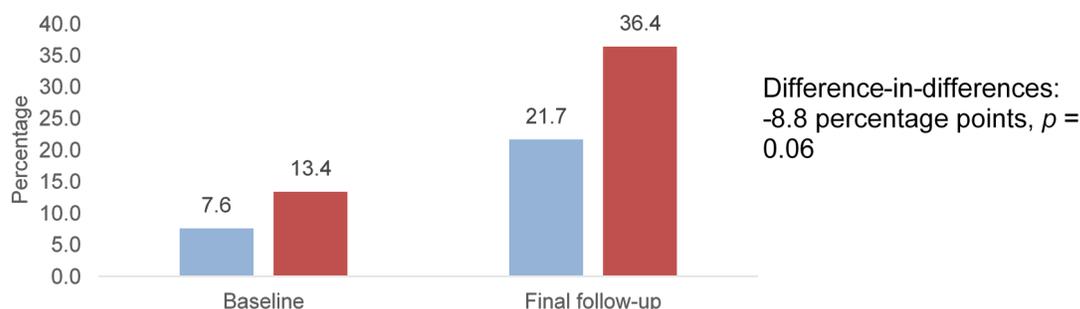
*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

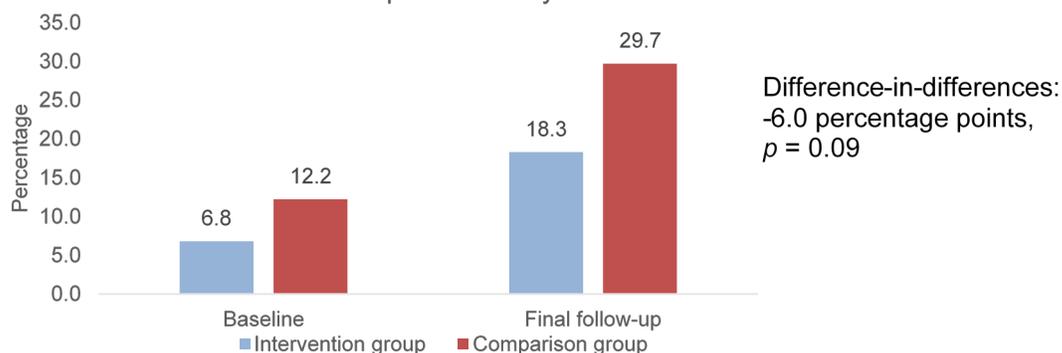
Students in New Jersey Teen PEP schools were also less likely to report engaging in unprotected sex in the three months before both the baseline and final follow-up surveys (Figure D.4). In the full sample of New Jersey students, 8 percent of students in New Jersey Teen PEP schools reported having had sex in the three months before the baseline survey, compared to 13 percent of students at comparison schools. The 6 percentage point difference grew to 15 percentage points by the time of the final follow-up survey. This implies a DD impact of -9 percentage points; however, the estimate is only marginally statistically significant ($p = 0.06$). Likewise, students in New Jersey intervention schools were 5 percentage points less likely at baseline than students in comparison schools to have recently had sex without any effective method of birth control. At the time of our final follow-up, this gap increased to 11 percentage points. However, the resulting DD impact estimate (-6 percentage points) was again only marginally significant ($p = 0.09$).

Figure D.4. Impacts of Teen PEP on unprotected sex, difference-in-differences in full New Jersey sample

Had sexual intercourse without a condom in the three months prior to survey^a



Had sexual intercourse without any effective method of birth control in the three months prior to survey



Source: Teen PEP baseline and final follow-up surveys, full sample.

Notes: The sample includes 242 intervention-group students and 96 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. Estimates are weighted to give each school equal weight. Estimates are in percentages. P -values are corrected for clustering at the school level. None of the estimates are significant at the 0.05 level.

^aConfirmatory outcome.

For both outcomes in the unprotected sex domain, the estimated impacts increase and become statistically significant when we use our most rigorous estimation method to produce impacts (Table D.4). Using propensity-score matching and regression, we estimate that 22 percent of students in New Jersey Teen PEP schools had sex without a condom in the three months before the 18-month follow-up survey, compared to 48 percent of students in New Jersey comparison schools. The 26 percentage point impact is statistically significant ($p < 0.01$). Likewise, this estimation approach implies that Teen PEP reduced rates of sex without any effective method of birth control by 22 percentage points ($p = 0.01$).

Table D.4. Impacts of Teen PEP on unprotected sex estimated by regression analysis, New Jersey sample

Variable	Intervention group mean	Comparison group mean	Difference	p-value
Had sex without a condom in the past three months ^a				
Propensity-score matched sample	22.2	48.2	-26.0	<0.01**
Propensity-score trimmed sample	22.1	36.9	-14.8	0.02*
Full sample	21.3	37.0	-15.7	0.01**
Had sex without any effective birth control method in past three months				
Propensity-score matched sample	18.8	41.0	-22.2	0.01**
Propensity-score trimmed sample	18.2	31.0	-12.8	0.05*
Full sample	17.8	30.5	-12.7	0.03*

Source: Teen PEP follow-up survey.

Notes: Impact estimates and group means are regression-adjusted to account for baseline characteristics (see Appendix B for a full list of covariates). Estimates are weighted to give each school equal weight (full and propensity-score trimmed samples) or each intervention school equal weight (propensity-score matched sample). Estimates are in percentages. The propensity-score matched sample included 217 intervention-group students and 55 comparison-group students. The propensity-score trimmed sample included 217 intervention-group students and 95 comparison-group students. The full sample included 242 intervention-group students and 96 comparison-group students. Item-specific nonresponse limits sample size for some outcomes. *P*-values are corrected for clustering at the school level.

^aConfirmatory outcome.

*Significantly different from zero at the 0.05 level, two-tailed test.

**Significantly different from zero at the 0.01 level, two-tailed test.

Impacts estimated using regression analysis on the full or propensity-score trimmed samples of New Jersey students were also both sizable and statistically significant (Table D.4). These methods imply that Teen PEP was associated with a decrease in the rate of sex without a condom of between 15 and 16 percentage points ($p < 0.02$) and a decrease in the rate of sex without any effective form of birth control of 13 percentage points ($p < 0.05$).



PPA
Evaluation of Adolescent
Pregnancy Prevention Approaches

<http://www.hhs.gov/ash/oah/>

