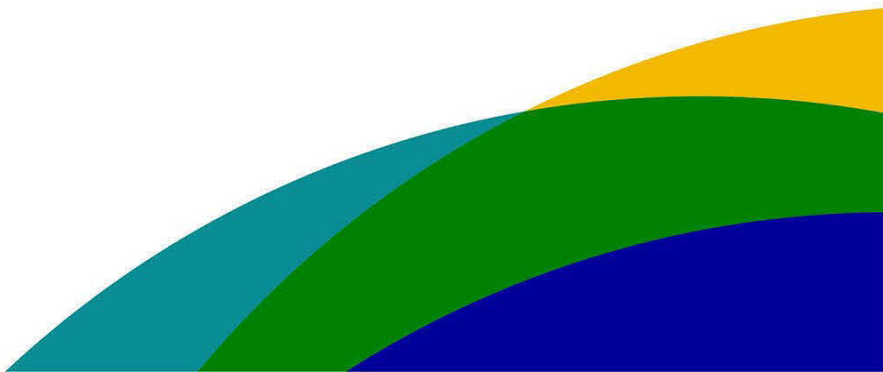
OASH | Office of Population Affairs

Evaluation Report:

THE IMPACT OF THE MAKING PROUD CHOICES! TEEN PREGNANCY PREVENTION CURRICULUM

May 2022



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

Purpose Statement

In this report, we share the findings from an impact and implementation study of Making Proud Choices! (MPC), a teen pregnancy prevention curriculum. In 2015, Mathematica was awarded a contract by the Office of Adolescent Health (now the Office of Population Affairs), U.S. Department of Health and Human Services (HHS), to add to the evidence base on teen pregnancy prevention (TPP) programming. At that time, MPC was one of the most widely implemented evidence-based TPP programs nationwide; however, only a single study, conducted by the developers of MPC more than 20 years ago, showed the effectiveness of the program (Jemmott et al. 1998). The program version, the participants' ages, and the implementation setting studied in the original evaluation were substantively different from the current version of the program, the populations being served, and the implementation settings most common in the field today. This new study, based on a large, rigorous random assignment evaluation, provides much-needed evidence on the effectiveness of MPC as it is currently implemented.

Authors

Russell Cole, Mathematica

Theresa Schulte Neelan, Mathematica

Andrew Langan, Mathematica

Betsy Keating, Mathematica

Jennifer Walzer, Mathematica

Subuhi Asheer, Mathematica

Susan Zief, Mathematica

Submitted to

Dr. Amy Farb, Evaluation Specialist, Office of Population Affairs, Office of the Assistant Secretary for Health, U.S. Department of Health and Human Services

Suggested citation: Cole, R., T. Schulte Neelan, A. Langan, B. Keating, J. Walzer, S. Asheer, and S. Zief. "The Impact of the Making Proud Choices! Teen Pregnancy Prevention Curriculum." Washington, DC: Office of Population Affairs, Office of the Assistant Secretary for Health, U.S. Department of Health and Human Services, 2022.

Prepared for OPA under contract number:
HHSP233201500035I/HHSP23337014T

Disclaimer

This publication was supported by Award No. HHSP233201500035I/HHSP23337014Tfrom the Office of Population Affairs (OPA). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of OPA or HHS.

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

Acknowledgements

Many organizations and people contributed in significant ways to this evaluation of the Making Proud Choices! (MPC) program. First, we acknowledge the invaluable support and guidance from the Office of Population Affairs (OPA, formerly the Office of Adolescent Health) in the U.S. Department of Health and Human Services (HHS), and especially Amy Farb, the OPA project officer for the impact evaluation.

We are thankful for the staff at the four organizations who provided MPC to youth so effectively: the Mobile County Health Department, Mobile, Alabama; the YMCA of Detroit, Michigan; Planned Parenthood of Southwest Ohio; and Better Family Life of St. Louis, Missouri. We also appreciate the assistance of the school and district staff and administrators who allowed us into their schools for the study and are grateful to the parents for granting permission, and to youth for actively participating in data collection.

We thank Alexandra Eisler and Valerie Sedivy at the Healthy Teen Network for their guidance on the MPC program, skill in holding trainings, and effective technical assistance. At Decision Information Resources, we thank Natalia Ibanez, Stephanie Lane, and Carol Pistorino, who oversaw the data collection on fidelity and attendance logs and observations. At Mathematica, we thank Melissa Thomas, Jennah Gosciak, Charles Tilley, John Deke, and Dan Thal, who contributed their survey, programming, data collection, and data analysis expertise; and Jean Knab, whose comments improved the quality of our report. Aunsara Dennis and Gwyneth Olsen created the report graphics, Molly and Jim Cameron edited the final report, and Dorothy Bellow produced it. Finally, we thank Matthew Chinman, Kelleen Kaye, and David Wolfe for their feedback on a draft version of this report.

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

Contents

Exhibits

Executive Summary

There is a large and growing evidence base showing the effectiveness of teen pregnancy prevention (TPP) programs, yet most have evidence from only a single small-scale efficacy trial (Goesling 2015). Making Proud Choices! (MPC) is an extremely popular evidence-based program that has been implemented in a large number of settings nationwide (Murphy et al. 2021). Despite its popularity, only a single study shows its effectiveness, and the evidence from its evaluation is more than 20 years old (Jemmott et al. 1998). To provide evidence of the effectiveness of MPC as implemented with youth today, the Office of Population Affairs commissioned Mathematica and its subcontractors, Decision Information Resources and the Healthy Teen Network, to conduct a rigorous national evaluation of the program.

The MPC program aims to provide youth with the information and tools they need to reduce their risk of sexually transmitted infections (STIs), human immunodeficiency virus (HIV), and pregnancy. The curriculum emphasizes abstinence as the safest choice for avoiding pregnancy and STIs, but also encourages youth to use condoms and other methods of birth control if they do have sex. The school-based version of the program tested in this evaluation includes fourteen 40-minute lessons. The logic model for the MPC program suggests that the program first affects risk and protective factors for risky sexual behavior, including knowledge, attitudes, beliefs, and skill and self-efficacy in using condoms and refusing to have sex. Improvement in these proximal outcomes is expected subsequently to influence sexual behaviors and, eventually, the health goals of the MPC program: reducing teen pregnancy and STIs.

This evaluation used a cluster randomized controlled trial design to assess MPC's effectiveness. High schools within selected cities were randomly assigned to one of two conditions: (1) MPC implemented by trained health educators or (2) business as usual. Before random assignment, each school was able to select the classroom setting for the study (where MPC would be delivered were the schools randomized to receive it, or where business-as-usual programming would occur); schools most often targeted health classes. The study took place in 15 schools across four cities where the teen birth rate and/or STI rate was markedly higher than the national average, and schools were re-randomized to condition up to three times, depending on how many years the school participated. Altogether, the study included 31 randomized clusters and more than 2,800 youth, most of whom were in 9th and 10th grades.

The evaluation relied on self-reported youth survey data and qualitative and quantitative implementation data. Youth in the study completed a baseline survey administered before programming began and a follow-up outcome survey administered approximately six months after the end of programming—nine months after baseline, on average. The surveys included adapted versions of items previously used in the random assignment evaluation of MPC and other survey items adapted from similar federal evaluations or developed for the study. The surveys measured antecedents to sexual behavior (risk and protective factors), sexual behaviors, and MPC's health goals (prevention of STIs and teen pregnancies). Program implementation data came from fidelity and attendance logs, observations, interviews, study youth focus groups, staff surveys, and technical assistance logs.

The survey and implementation data suggest a strong difference in experiences across the two conditions examined in this study. A combination of qualitative and quantitative implementation data suggest that health educators delivered the program as intended with high quality, and that the majority of youth received a large dose of the program. In addition, youth in the control group reported receiving far less information about teen pregnancy prevention and sexual health during classes and interactions with reproductive health care providers than youth in the MPC group.

The evaluation showed several large, statistically significant, and favorable impacts on nearly all of the risk and protective factors for risky sexual behavior—the outcomes most proximal to the content of MPC. The MPC program significantly improved knowledge of HIV/STIs, pregnancy, condoms, and other contraceptives; attitudes and beliefs about condoms; and self-efficacy in using condoms, negotiating condom use, and refusing sex.

Among the 10 outcomes that measured sexual behaviors (for example, sexual initiation, sex without a condom) and MPC health goals (that is, pregnancy and STI prevention), the study observed one statistically significant impact: the MPC group reported significantly fewer episodes of sex in the past three months relative to the control group. The magnitude of the impact estimates for other behavioral and MPC health goal outcomes were small and nonsignificant, but the direction of most findings suggested a favorable effect of MPC.

The evaluation also examined MPC's effectiveness across a variety of subgroups to better understand the extent to which its effect varied across sites or youth. The original study of MPC (Jemmott et al. 1998) showed that the program was particularly effective at improving sexual behavior outcomes among the subset of youth who were sexually active at baseline. Although the magnitude of the impact estimates suggest MPC is more effective for youth who were sexually experienced at baseline, there were no statistically significant impacts observed on any sexual behavior or MPC health goal outcomes for this subgroup or any other subgroups examined in this study.

Favorable and statistically significant impacts of MPC observed in this study:

Risk and protective factors

Knowledge of:

- HIV/STIs
- Pregnancy
- Condoms
- Other forms of contraception

Beliefs about:

- Condom use

Attitudes about:

- Condom use

Skill and self-efficacy related to:

- Condom use
- Condom use negotiation
- Refusal of sex

Sexual behaviors

- Frequency of recent sex

The finding that MPC generally produced larger and more significant impacts on the risk and protective factors than the more distal behavioral outcomes is consistent with the logic model for MPC. The proximal outcomes—those related to knowledge, beliefs, attitudes, and skills and self-efficacy—aligned well with the content and activities provided by MPC and were expected to change the most at the short-term six-month follow-up assessment. The favorable and significant impacts on risk and protective factors may have only begun to manifest themselves as significant impacts on behavior.

Overall, the MPC evaluation found favorable program effects, including the type of findings that TPP evidence review criteria can use to characterize MPC as a program with evidence of effectiveness. The program showed consistently favorable, large, and statistically significant impacts on the outcomes identified in the logic model as proximal short-term outcomes. Importantly, for behavioral outcomes, it showed one significant favorable impact, several nonsignificant impacts, and no significant unfavorable impacts. Ideally, the study would have produced long-term follow-up data to understand whether the impacts on risk and protective factors subsequently produced significant favorable changes on more of the sexual behaviors and, eventually, on MPC health goals as hypothesized by the logic model. Thus, more research into the longer-term effectiveness of MPC may be needed to address this remaining gap in the literature. However, even without these longer-term findings, the current study provides evidence to suggest that as implemented today, MPC should continue to be considered an evidence-based program that favorably affects behavioral outcomes.

I. Introduction

The federal emphasis on identifying and using evidence-based approaches to teen pregnancy prevention (TPP) began in 2010 with congressional funding for the competitive TPP grant program to be administered by the Office of Adolescent Health, which in 2019 was incorporated into the Office of Population Affairs (OPA), also within the Office of the Assistant Secretary for Health (Kappeler and Farb 2014). The TPP program was one of six early evidence-based initiatives proposed by the Obama administration and authorized by Congress to increase the use of data and evidence in social policy (Haskins and Margolis 2015). This competitive grant program provides roughly $100 million annually to state and local organizations to implement evidence-based and promising new TPP programs.

The TPP evidence review was another federal initiative focused on building evidence on TPP. This initiative was jointly sponsored by three agencies within the U.S. Department of Health and Human Services: the Office of the Assistant Secretary for Planning and Evaluation, the Family and Youth Services Bureau of the Administration for Children and Families (ACF), and OPA. The TPP evidence review involved a systematic review of the research literature to identify TPP programs effective at reducing risky sexual behaviors. One of the key early findings from this review was that most TPP programs have been evaluated only once—typically in small-scale "efficacy" trials conducted in closely managed settings, often by the program developer (Goesling et al. 2014). Although these efficacy trials are important for establishing initial evidence of a program's success, it is equally important to know how programs perform when implemented on a broader scale, with different populations, or in new settings (Valentine et al. 2011).

To understand the extent to which the original evidence of program effectiveness was robust in new settings and with new populations, OPA funded studies of TPP programs that had existing evidence of effectiveness. OPA funded more than a dozen of these replication studies with the first cohort of TPP program grantees; most were led by individual grantees in partnership with an independent evaluator (Farb and Margolis 2016). A smaller number of such studies were conducted as part of the federal TPP Replication Study led by OPA and the Office of the Assistant Secretary for Planning and Evaluation (Abt Associates 2015) and the Personal Responsibility Education Program (PREP) Multi-Component Evaluation led by the Office of Planning, Research, and Evaluation and the Family and Youth Services Bureau within ACF (Goesling et al. 2018). Several other programs have also been evaluated by academic researchers outside of the current set of federally funded replication studies (Coyle et al. 2013; Jemmott et al. 2010; Markham et al. 2012, 2014).

In 2015, Mathematica was awarded a contract from OPA to further improve the evidence base on TPP programming, which included identifying additional evidence-based programs in need of replication research. At the time, Making Proud Choices! (MPC) was one of the most widely implemented evidence-based TPP programs among federally funded grantees nationwide. More than 30 grantees that OPA funded in 2010 and 2015 were implementing MPC. MPC was also widely implemented through PREP, with nearly 70 providers implementing the program. Despite MPC's popularity, only one study showed the effectiveness of this program, which its developers conducted more than 20 years ago (Jemmott et al. 1998). Importantly, the program version, age of study participants, and implementation setting examined in this original evaluation were substantively different from the current program version, populations being served, and implementation settings most common in the field today. Therefore, a large number of grantees were implementing a program with evidence of effectiveness, but the evidence was not based on the current, typical implementation of MPC most prevalent in the field. To update the evidence about the effectiveness of this popular program, OPA issued a request for proposals for a rigorous impact and implementation evaluation and awarded the contract to conduct the national evaluation to Mathematica.

A. Making Proud Choices! (MPC)

The MPC curriculum aims to provide youth with the information and tools they need to reduce their risk of sexually transmitted infections (STIs), human immunodeficiency virus (HIV), and pregnancy. The curriculum emphasizes abstinence as the safest choice for avoiding pregnancy and STIs, but also encourages youth to use condoms and other methods of birth control if they do have sex. The curriculum includes video clips, youth role-playing activities, and group discussions. It also involves skill-building activities, including opportunities for youth to practice correct condom use, refusal techniques, and safer-sex negotiation. Additional detail about the program is in Section II.

B. The current study

This evaluation used a cluster randomized controlled trial (RCT) design to assess the effectiveness of MPC. High schools within cities were randomly assigned to one of two conditions: (1) MPC implemented by health educators provided by local health education organizations or (2) business as usual. Before random assignment, each school was able to select the classroom setting for the study (where MPC would be delivered if the school was randomized to receive it or where business-as-usual programming would occur); most often, schools targeted health classes.

The study was implemented over the course of three school years (2016−2017, 2017−2018, and 2018−2019), with youth in each year considered a separate evaluation cohort and participating schools and new cohorts of eligible youth randomized to condition each year. The study took place in 15 schools across four cities where the teen birth rate and/or STI rate was markedly higher than the national average, and schools were re-randomized to condition up to three times, depending on how many years the school participated. Altogether, the study included 31 randomized clusters and more than 2,800 youth, most of whom were in 9th and 10th grades.

Youth in the study completed two waves of self-report surveys. They completed a baseline survey administered before programming began and a follow-up outcome survey approximately six months after programming ended (approximately nine months after baseline, on average). The surveys included adapted versions of items used in the prior evaluation of MPC and other survey items either adapted from similar federal evaluations or developed for the study. The items included measures of antecedents to sexual behavior (knowledge, attitudes, beliefs, self-efficacy), sexual behaviors, and incidence of STIs and teen pregnancies—all expected outcomes per the MPC logic model (described in Section II). Program implementation data included fidelity and attendance logs, observations, interviews, study youth focus groups, staff surveys, and documentation of trainings and technical assistance (TA) provided to the health educators.

The impact study was designed to measure the effectiveness of MPC on outcomes aligned with the program logic model. This study provides information on the full study sample, as well as the extent to which impacts varied for subgroups, such as by gender and sexual experience levels at baseline. Implementation data provide insight into the extent to which the program was delivered with fidelity and quality, and the experiences of health educators and youth in the study. The study also provides data necessary to assess the strength of the difference between the experiences of the MPC and control groups in the study.

C. Road map

The remainder of this report includes four sections, along with additional information in the appendices. Section II describes the MPC program. Section III describes the impact study design, data sources, and an outline of the analytic methods. Section IV presents the implementation findings, and Section V presents the impact findings. Section VI concludes the report. Appendix A presents additional information on the data, methods, and analyses for both the impact and implementation studies. Appendix B presents

detailed information on baseline equivalence, and Appendix C presents detailed subgroup findings and sensitivity analyses.

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

II. Making Proud Choices! Program Description

In 2015, when OPA requested an evaluation of MPC, it was one of the most widely implemented evidence-based TPP programs among federally funded grantees nationwide. Since 2010, it has been implemented by more than 30 of OPA's TPP grantees and was one of four most commonly implemented programs under ACF's PREP program (Murphy et al. 2021). The popularity of this program, coupled with the lack of recent rigorous evidence of its effectiveness, suggested a need for a federal evaluation to understand its effectiveness as it is commonly implemented today.

A. Previous research on MPC

There is one study of MPC that met TPP evidence review standards and provides favorable evidence of effectiveness on sexual behavioral outcomes (Jemmott et al.1998). This study used an individual-level random assignment design to place youth into (1) a safer-sex program (later titled MPC), (2) an abstinence program (later titled Making a Difference!), or (3) a control group receiving information on cardiovascular disease and cancer. Each group met for four-hour sessions on two consecutive Saturdays. The study took place in the mid-1990s in a middle school in a low-income area of Philadelphia and was conducted with African American students in grades 6 and 7. This population was targeted because they were at particularly high risk of contracting HIV during the acquired immunodeficiency syndrome (AIDS) crisis occurring at that time (Centers for Disease Control and Prevention 1996). The adult facilitators for this program had experience in providing programming to the target population, and the original MPC developer oversaw program implementation.

The TPP evidence review

From 2009 to 2017, the U.S. Department of Health and Human Services sponsored a systematic review of TPP research literature to identify programs with evidence of effectiveness in reducing teen pregnancies, STIs, and sexual risk behaviors. The review developed evidence standards that could be used to categorize the rigor and credibility of evidence of the effectiveness of TPP research literature.

The study examined impacts at three, six, and 12 months after the program and revealed several favorable and statistically significant impacts on behavioral outcomes. Three months after the program, the MPC group revealed significantly lower prevalence of sex without a condom (a difference of 8 percentage points) relative to the control group and a significantly lower frequency of sex events without a condom (0.14 fewer days of sex without condoms in the past three months). These effects were even more substantial for the subset of youth who were sexually experienced at baseline (30 percentage points and 0.59 fewer days, respectively). At both the six- and 12-month follow-ups, the following outcomes also revealed statistically significant and favorable program impacts for the subset of sexually active youth at baseline: less frequent sexual intercourse, more frequent condom use, and less frequent sex without a condom.

B. Description of MPC

The MPC curriculum is designed to give youth the information and tools they need to reduce their risk of STIs, HIV, and pregnancy. The curriculum emphasizes abstinence as the safest choice for avoiding pregnancy and STIs, but also encourages youth to use condoms and other methods of birth control if they do have sex. The curriculum includes video clips, youth role-playing activities, and group discussions. It also involves skill-building activities, including opportunities for youth to practice these skills, for correct condom use, refusal techniques, and safer-sex negotiation.

The MPC program has been revised several times in the past two decades; the version evaluated in this study is the fifth edition. The overall messages and goals of the curriculum have not changed since the original version studied in Jemmott et al. (1998), but videos and other activities have been updated

slightly over time. For instance, the fifth edition of MPC includes updated information on birth control methods, presented in order of effectiveness, with a greater emphasis on long-acting reversible contraceptives than earlier versions. The curriculum now incorporates trauma-informed approaches and inclusivity and sensitivity toward LGBTQI+ youth. Unlike some prior versions, the fifth edition uses the original ordering of program lessons, which introduced HIV before other topics, such as STIs and pregnancy prevention.[1]

At the time the evaluation was designed, there were three versions of the fifth edition MPC curriculum, each intended to be used in different implementation settings or with different populations: Original, School, and Out-of-Home. The Original Edition contained eight 1-hour lessons intended for implementation in community settings. The School Edition contained fourteen 40-minute lessons intended for implementation in classroom settings where a one-hour lesson is typically less feasible. The Out-of-Home Edition included ten 75-minute lessons intended for implementation in foster homes, independent and transitional living facilities, and juvenile justice settings.

Findings from the PREP Design and Implementation Study indicated that the majority of MPC programs were being implemented in high schools (Zief et al. 2013). A review of the 2015 OPA grant applications also suggested that MPC was most often selected for school-based settings. Therefore, the evaluation selected the MPC School Edition—the version of the fifth edition of MPC designed for school classroom settings—for the study (see Exhibit II.1 for a list of the MPC School Edition lessons).

Exhibit II.1. MPC School Edition lessons

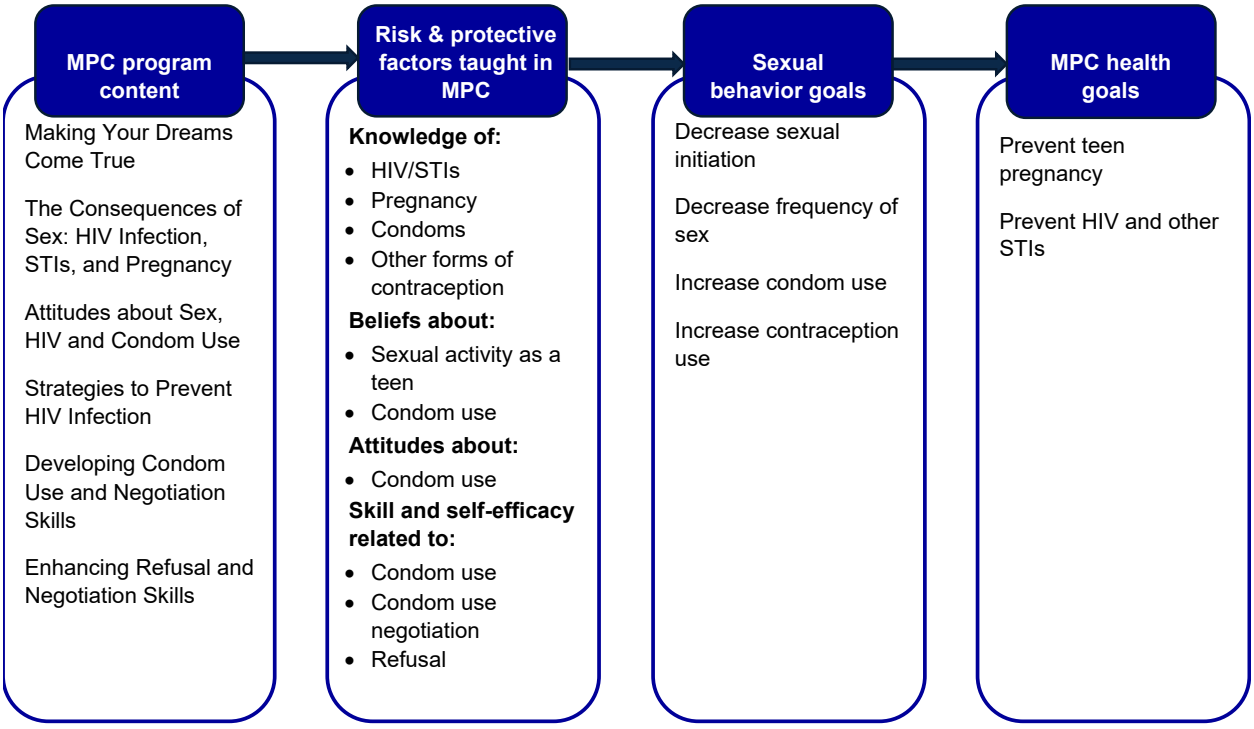
Lesson number	Lesson title
1	Getting to Know You and Steps to Making Your Dreams Come True
2	The Consequences of Sex: HIV Infection: Part 1
3	The Consequences of Sex: HIV Infection: Part 2
4	Attitudes about Sex, HIV, and Condom Use
5	Strategies to Prevent HIV Infection: Stop, Think, and Act: Part 1
6	Strategies to Prevent HIV Infection: Stop, Think, and Act: Part 2
7	The Consequences of Sex: STIs: Part 1
8	The Consequences of Sex: STIs: Part 2
9	The Consequences of Sex: Pregnancy: Part 1
10	The Consequences of Sex: Pregnancy: Part 2
11	Developing Condom Use and Negotiation Skills: Part 1
12	Developing Condom Use and Negotiation Skills: Part 2
13	Enhancing Refusal and Negotiation Skills: Part 1
14	Enhancing Refusal and Negotiation Skills: Part 2

The logic model for the MPC evaluation shows the expected sequence of how the program lessons influence (1) risk and protective factors, (2) sexual behaviors, and ultimately (3) MPC's health goals (Exhibit II.2). Improvement in four risk and protective factors (knowledge, beliefs, attitudes, and skill and self-efficacy) are expected to influence sexual behavior outcomes (sexual initiation, frequency of sex, and condom and contraceptive use). Reductions in these risky sexual behavioral outcomes are expected to achieve MPC's health goals of reducing teen pregnancy and HIV and other STIs.

[1] See http://www.etr.org/ebi/assets/File/2016-Updates-MPC-5thEd.pdf for a more detailed, bulleted summary of the updates in the fifth edition of MPC.

The logic model helps to frame the expected sequence of when different types of impacts might be observed in this evaluation. The risk and protective factors were expected to be the outcomes most likely to change in a short-term follow-up assessment. Impacts on sexual behaviors and MPC health goals were expected to occur at a long-term follow-up assessment once youth had enough time to change their behavior in response to changes in their risk and protective factors. As noted in the introduction, this evaluation was able to measure changes in the short term—approximately six months after the end of programming.

Exhibit II.2. Logic model for the MPC evaluation



C. Delivery of MPC: health educators or school teachers

For school-based delivery of MPC, either trained health educators from an organization partnering with the school or classroom teachers can deliver the program. A benefit of using external health educators as facilitators is that they often are knowledgeable about HIV/AIDS, STIs, and youth sexuality, and have experience in implementing evidence-based programs. Given their familiarity with the content and delivery of similar programs, health educators should be expected to implement the MPC curriculum with fidelity, provided they have received sufficient training. Furthermore, youth may also be more comfortable when receiving and discussing sensitive information from an outside health educator rather than the classroom teacher with whom they interact on a daily basis (Pound et al. 2016). On the other hand, a key benefit of training teachers to deliver MPC is the potential sustainability of the programming in schools. Once trained, the teachers can continue to implement MPC during their careers at essentially no cost to the school (aside from the purchase of additional materials as needed). For the purpose of this study, which included re-randomization of schools across each of three years of implementation, the study team brought trained health educators into the schools instead of training classroom teachers so the individuals delivering the program could move across schools based on the results of random assignment.

D MPC training and technical assistance

At the time this evaluation of MPC was designed, ETR, the distributor of MPC, recommended that all educators implementing the program receive training on the curriculum to prepare them to implement it with fidelity. ETR offered a two-day in-person training with follow-up support as needed. [2] In addition, ETR offered TA before, during, and after program implementation to support high quality implementation and address questions about adaptations.

For this study, Mathematica selected the Healthy Teen Network (HTN) as the MPC training and TA provider and required all facilitators to participate in this training and TA. Mathematica chose HTN for this role based on its extensive experience in training TPP grantees on MPC and other evidence-based programs. Before implementing MPC for the study, health educators participated in a two-day in-person training on the program. These educators first completed individual online activities on the theory and background of MPC and then attended interactive sessions in which a certified HTN trainer explained each lesson of the curriculum and modeled best practices for implementation with youth. Following this initial training, HTN provided support to health educators over the phone throughout implementation to ensure they felt prepared, help them troubleshoot issues, and discuss best practices for implementing the program with youth while adhering to the scripted curriculum.

[2] In 2020, ETR started offering virtual training, which includes three 2.5-hour interactive sessions delivered over three weeks. See details at https://www.etr.org/ebi/programs/making-proud-choices/.

III. Design, Research Questions, Data Sources, Outcome Measures, and Analysis Methods

The MPC impact evaluation used a random assignment design to estimate the program's effectiveness in changing youth outcomes. This section briefly describes the study design and approach for measuring and describing implementation and program effectiveness; more details are in Appendix A.

A. Research questions

The impact evaluation of MPC was intended to answer several research questions about the implementation of and effectiveness of this program.

1. Implementation research questions

The study team answered the implementation research questions using several sources of qualitative and quantitative data, including fidelity and attendance logs, independent observations, youth and educator surveys, interviews with health educators and school staff, and focus groups with youth. The implementation study addressed the following research questions:

1. **MPC implementation:** How was MPC delivered and what were the experiences of health educators and youth in the study?
    a. What training and TA did health educators receive to support program delivery?
    b. What proportion of MPC lessons did youth attend, on average?
    c. To what extent did health educators deliver MPC with fidelity and quality?
    d. How different were the experiences of youth receiving MPC relative to the business-as-usual control group?

2. Effectiveness research questions

The study team answered the effectiveness research questions using survey data collected roughly six months after program delivery (approximately nine months after baseline).[3] The team estimated impacts for all of the outcomes shown in the logic model in Exhibit II.2:

1. **Impacts on risk and protective factors:** What was the effect of MPC on the following risk and protective factors that are antecedents to sexual behavior?
    a. Knowledge of HIV/STIs, pregnancy, condoms, and other forms of contraception
    b. Beliefs about sexual activity as a teen, communication with partners, and condom use
    c. Attitudes about condom use
    d. Skill and self-efficacy related to condom use, condom use negotiation, and refusal

[3] The original plan for the MPC impact evaluation was to focus only on risk and protective factor outcomes in the short-term follow-up assessment and focus on sexual behavior and MPC health goal outcomes in the longer-term follow-up. This analytic approach was aligned with the logic model and theory of change for MPC, in which changes in risk and protective factors would promote subsequent behavioral changes. However, given that the longer-term follow-up data collection was not funded, this report looks at all three domains of outcomes by using short-term follow-up survey data.

2. **Impacts on sexual behaviors:** What was the effect of MPC on the following sexual behavior outcomes?
    a. Sexual initiation
    b. Frequency of sex
    c. Having sex without a condom
    d. Having vaginal sex without birth control
3. **Impacts on MPC health goals:** What was the effect of MPC on the incidence of the following longer-term health outcomes?
    a. Pregnancy
    b. STIs (including HIV)
4. **Impacts for subgroups:** Was the effect of MPC the same across cities, genders, instances of random assignment (see details about the assignment process below), or sexual experiences measured at baseline?

B. Recruitment

Beginning in spring 2016, the study team worked with OPA to identify partner organizations to participate in the study. The team explored a variety of candidate provider organizations that had experience in providing MPC in local school districts. The team prioritized organizations working in urban areas where the teen birth or STI rates were markedly higher than state or national averages to ensure the study was serving a high-risk population. Ultimately, the study team recruited implementation partner organizations with experienced health educators in four cities/metropolitan areas: (1) Mobile, Alabama; (2) Detroit, Michigan; (3) Cincinnati, Ohio; and (4) St. Louis, Missouri.[4] The team then worked with the partner organizations, which were all well known by the local school districts in these cities, to recruit districts and schools for the study.

C. Study design

This evaluation used a cluster RCT design to assess the effectiveness of MPC, with high schools within cities randomly assigned to one of two conditions: (1) MPC implemented by health educators or (2) business as usual. Health educators from local partner organizations received training and TA from HTN and implemented the program.

The study was implemented over the course of three school years, with youth in each year considered a separate evaluation cohort. Before random assignment, each school was able to select the classroom setting for the study (where MPC would be delivered if the school were randomized to receive it or where business-as-usual programming would occur); most often, schools targeted health classes. [5] Cohort 1 included students from six schools in one city in spring 2017. Cohort 2 included students in 10 study schools in four cities during the 2017–2018 school year, and Cohort 3 included students in 15 schools and four cities during the 2018–2019 school year. Before the start of each cohort, Mathematica conducted random assignment. Half of the participating schools in a participating city were randomly assigned to receive the MPC program provided by a health educator and half were assigned to continue their business-as-usual class.

[4] For brevity, the remainder of this report refers to all four locations as “cities,” rather than “cities/metropolitan areas.”

[5] As a result of this design feature, the services available or provided to the comparison group were different at each school. However, the comparison programming never included a competing evidence-based program intended to affect sexual behavior outcomes.

Several of the schools participated in the study across multiple cohorts. These schools were re-randomized to condition each year they participated. As a result, many schools switched conditions across cohorts. (For example, a school received MPC in Cohort 1 but then did not receive it in Cohort 2.) This re-randomization was feasible because (1) different students were eligible for participation in the evaluation each year (see the study consent process subsection that follows for details), and (2) the health educators could move across schools between cohorts (that is, the treatment was impermanent within a school). In total, the study team randomly assigned 31 school clusters containing roughly 2,800 youth to condition in the evaluation.

D. Data sources

This analysis is based on two waves of survey and program implementation data from each year of programming. Program implementation data included fidelity and attendance logs, observations, interviews, study youth focus groups, staff surveys, and TA logs.

1. Survey data

Youth in the study completed two waves of self-report surveys—a baseline survey and a follow-up survey. The timing of data collection was determined by the schedule for programming in the treatment schools, and students in the treatment and control schools were surveyed at the same time. Youth completed the baseline survey shortly after parental consent and before programming began in the treatment schools. They also completed a follow-up survey approximately six months after programming ended in the treatment schools (roughly nine months after baseline, on average).[6] The surveys included adapted versions of items previously used in the random assignment evaluation of MPC (Jemmott et al. 1998) and other survey items adapted from other federal evaluations of similar programs or developed for the study to address outcomes in the logic model. They included questions on demographics, knowledge, attitudes, beliefs, self-efficacy, sexual behaviors, incidence of STIs and teen pregnancies, and exposure to information one would expect to receive in a teen pregnancy prevention program.

2. Implementation data

Data on program implementation came from several sources:

- **Fidelity and attendance logs.** Each health educator completed an online fidelity and attendance log after each MPC lesson. They recorded attendance data for all students in the class who consented to participate in the study. They also answered questions about whether they fully completed all activities or made any changes or additions, and shared information on any challenges they encountered.
- **Observations of implementation and health educator training.** The study team completed observations of about 5 percent of all MPC lessons implemented for the study. The team purposively selected lessons for observations, prioritizing key lessons identified by the developer and HTN as critically important for improving behavioral outcomes (Lessons 8 and 11 through 14).[7] Observers assessed several metrics of implementation quality, including health educator comfort and preparedness, and youth engagement. Separately, the study team observed two of the four MPC trainings conducted by HTN to capture data on training content and methods, and health educator engagement and comfort with the curriculum.

[6] The timing of follow-up data collection varied across participating schools, with the expectation that it would occur roughly six months after programming was offered. If program implementation occurred in the fall semester, follow-up data collection was scheduled to occur in the spring semester. Similarly, if program implementation occurred in the spring semester, follow-up data collection was scheduled to occur in the fall of the next school year.

[7] In an initial meeting during the design phase of the study, Loretta Jemmott, the developer of MPC, indicated that the condom demonstration lesson and practice were critical features of the program for improving behavioral outcomes.

- **Summaries of TA provided by HTN.** The HTN TA providers completed logs to document the TA they provided to the health educators implementing MPC. These logs noted the contact type (email or phone), health educator(s) involved, issues discussed, and resolutions or next steps identified.
- **Interviews.** The study team conducted interviews with health educators to better understand their experiences in implementing MPC, including challenges and lessons learned. The team also conducted interviews with school staff to gather additional information on program implementation and the experiences of the control group. Respondents included principals, guidance counselors or social workers, and the teachers whose classes were in the evaluation.
- **Student focus groups.** The study team spoke to 12 groups of students at treatment schools to understand their perceptions of the MPC program. The team conducted these focus groups in person during or immediately after students' participation in MPC.
- **Health educator survey.** Most of the health educators implementing MPC completed a survey on their background and experiences with training and program delivery. Health educators completed this survey once after completing their first implementation cohort.
- **Student surveys.** The study team also drew on data from the student follow-up surveys to obtain information on the teen pregnancy and sexual health programming youth received.

E. Study consent process

Several weeks before the start of programming, the study team coordinated with staff in all treatment and control schools to distribute and collect parental consent forms. The team used the same process for all schools, regardless of treatment condition. The team distributed parental consent forms to all students in the classrooms that the schools selected to target for the study; these forms asked parents to indicate whether or not their child could participate in the study. All youth whose names appeared on the initial rosters provided by the schools received the parental consent forms. Consent form collection continued until 90 percent of parents had returned them. Immediately before programming began in each school, the study team requested final rosters for the target class; students who remained on those rosters made up the eligible sample.[8]

In total, 1,010 (76.7 percent) youth in the treatment group and 1,128 (75.5 percent) youth in the control group provided consent to participate in the study. The study team assessed the potential threat of bias stemming from nonconsent. According to What Works Clearinghouse (2020) attrition standards, and using the "cautious" boundary, this study had *low* rates of attrition via nonconsent (23.9 percent overall attrition and a 1.2 percentage point differential due to nonconsent). The cautious boundary is used with studies when there is reason to believe that attrition may be more strongly related to the outcomes; this boundary was used previously as the attrition standard for the TPP evidence review (Mathematica 2014).

F. Baseline characteristics of the sample

A total of 2,035 youth completed baseline surveys—950 from the treatment group (94.1 percent of the consented sample) and 1,085 from the control group (96.2 percent of the consented sample).

The study sample was 15.6 years old on average, with 46 percent enrolled in 9th grade and 49 percent enrolled in 10th grade at the time of randomization. The student survey offered several categories in which to report gender; 46 percent identified their gender as male, 53 percent identified as female, and 1 percent self-identified in some other way. Demographically, 80 percent of students who reported their

[8] Students who entered target classes but whose parents did not provide consent for the study by the start of the program, either by not returning a consent form or returning it and denying consent, were not included in the evaluation sample. Furthermore, any students who joined the target class after the program began were never given a consent form, so they were also excluded from the sample.

race identified as Black only, 9 percent as White only, and 9 percent as two or more races. Only 4 percent of students identified their ethnicity as Hispanic or Latino.

At baseline, 34 percent of respondents reported having had sex at any time in the past, with 21 percent reporting a sexual encounter in the past three months. Approximately 66 percent of those youth who had a recent sexual encounter, about 14 percent of all respondents, reported having had sex without a condom in the past three months. At baseline, 2 percent of respondents indicated having been pregnant previously or gotten someone else pregnant, and 9 percent indicated they previously had been diagnosed with one or more STIs.

The study team found two statistically significant differences in characteristics between the treatment group and the control group (see Appendix B for details). There was a longer time between baseline and follow-up in the control group. This difference was due in part to variation in the follow-up data collection timing across sites (the data collection schedule for each site was determined before random assignment), even though original expectations were that this timing would be similar across sites as a result of random assignment. (See Exhibit A.2 in Appendix A for details on why the imbalance occurred.) In addition, there were slightly more youth in the control group who identified as being of two or more races. Importantly, some statistically significant differences would be expected as a result of random sampling error when examining nearly 50 baseline characteristics. In the study's benchmark analytic approach, the team adjusted for a large number of baseline characteristics, including the two that showed a significant baseline difference, and tested the robustness of the impact analysis by estimating impacts with a variety of different sets of covariates. (See more detail in the analysis methods subsection below.)

G. Outcome measures for the impact study

The study examined all outcomes described in the logic model for MPC (Exhibit II.2), including the short-term outcomes most proximal to the program, presented as the risk and protective factors antecedent to sexual behavior. The logic model also includes the more distal outcomes, including sexual behaviors and pregnancy and STI rates, presented as MPC health goals.

The study team created measures of risk and protective factors by combining survey items into reliable scales. These scales included measures of knowledge, beliefs, attitudes, skills, and self-efficacy. For the sexual behavior domain, the team used measures of sexual initiation, frequency of sex, risk for STIs, and risk for pregnancy—the measures of initiation and risky sex were intended to operationalize the program's emphasis on abstinence as the safest way to avoid pregnancy and STIs. Finally, to operationalize the MPC health goal domain, the team used indicators for pregnancy and testing positive for an STI.

See Appendix A for more details on the individual scales and the reliability of each outcome variable.

H. Follow-up data collection and response rates

Students completed a follow-up survey approximately six months after the end of programming—roughly nine months after baseline, on average. A total of 1,880 youth completed the follow-up survey—880 from the treatment group (87.1 percent of the consented sample) and 1,000 from the control group (88.7 percent of the consented sample). The study team again assessed the potential threat of bias stemming from both survey and item nonresponse among the survey respondents, and determined there was a *low* threat of nonresponse bias for all outcomes.

I. Analysis methods

1. Estimating impacts

The analysis sought to determine how MPC affected each of the outcome measures described earlier.[9] To estimate these impacts, the study team used a statistical model to compare the average outcomes in the treatment and control groups after accounting for several factors. These factors included (1) individual students' characteristics, (2) the possible relationship between outcomes for youth within the same school in a given cohort, and (3) the assignment of schools to condition within a given city or cohort. To increase the precision of the estimated treatment effects, all impact analyses in the study's main results benchmark model controlled for student characteristics, baseline risk behaviors, and survey timing, and a baseline measure of each outcome of interest.[10] To adjust for clustering, the team included school*cohort random effects and fixed effects for strata to account for the assignment of schools to treatment or control condition within a given city or cohort. For the main results, the team drew on data that excluded respondents with missing outcome responses and dropped inconsistent responses (such as individuals who responded at baseline that they had previously had sex but reported at follow-up that they had never had sex, or had had sex in the past three months but with zero partners).

To test the robustness of these results, the study team re-estimated all of its program impact analyses using several approaches. First it varied the respondent characteristics for which it adjusted, using models that eliminated the statistical controls for (1) randomization strata, (2) the length of time between baseline and follow-up and summer exposure indicators, and (3) all covariates in the model except the treatment indicator. The team also used three alternative approaches to data preparation in addition to analyzing the consistent set of survey responses among the complete case sample. One additional robustness check was similar to the benchmark approach but analyzed students' raw, unedited responses instead of dropping inconsistent survey responses. The second approach also retained inconsistent responses and used a method called complete case analysis, which excluded any respondents who lacked the full set of baseline covariates. The third approach filled in missing responses—both for baseline characteristics and outcome data—using a multiple imputation method.

The study team used two approaches to interpret evaluation results, beyond reporting the magnitude of the observed impact in raw and standardized effect size units. First, this study reports the statistical significance of impact estimates because it is familiar to many readers. In addition, for a TPP program to be labeled as evidence based under the TPP evidence review, it requires a statistically significant favorable finding on one or more behavioral outcomes. However, because statistical significance is often misinterpreted (Wasserstein and Lazar 2016; Greenland et al. 2016), the study also reports the probability that the program truly had a favorable (or unfavorable) impact. The study uses BAyeSian Interpretation of Estimates (BASIE; Deke and Finucane 2019) to assess the probability that MPC had a favorable impact on youth outcomes, given the findings from the present study and prior evidence regarding the range of impacts observed in past evaluation studies. This probability is called a Bayesian posterior probability. In the discussion of the study's findings, this report shares these probabilities and call attention to probabilities greater than 80 percent (suggesting a likely favorable impact of MPC) and less than 20 percent (suggesting a likely unfavorable impact of MPC).

To calculate the Bayesian posterior probabilities, the study team used available evidence appropriate for its domains of outcomes. For a summary of prior evidence on behavioral outcomes, the team drew on a

[9] The TPP evidence review focuses on well-defined individual behavioral outcomes, rather than aggregations of multiple measures. In addition, the TPP evidence review takes the presence of a single, statistically significant favorable behavioral outcome as evidence of program effectiveness, as long as there are no adverse, statistically significant impacts on behavior. As a result, the study estimates and reports impacts separately for each outcome of interest described in the MPC logic model.

[10] The covariates used in all models included the following: race/ethnicity, age at baseline, gender, parental presence, sexual orientation, school truancy, smoking, alcohol/drug use, sexual initiation, relationship status, length of time between surveys, and whether a summer vacation period fell between baseline and follow-up periods.

recent meta-analysis of TPP program evaluations conducted for OPA (Juras et al. 2019). Because this meta-analysis focused only on behavioral outcomes, for the risk and protective factor outcomes the study drew on a review of moderate and high-quality studies summarized in the What Works Clearinghouse (WWC), which covered a broad range of programs (Herrmann et al., 2019). The WWC reviews evidence of educational programs; often these studies focus on researcher-developed tests well aligned with the educational content being evaluated. That is, the outcomes on which the WWC typically reports and reviews are analogous to the risk and protective factors described as proximal outcomes for MPC. Aligning with expectations, the prior evidence suggests that the magnitude of impacts for proximal outcomes (from the WWC) are larger than those for distal behavioral outcomes from the meta-analysis. Given that this study reports Bayesian posterior probability estimates, which are less sensitive than *p*-values to testing multiple outcomes within a single domain, the study team did not make a multiple comparison adjustment to its *p*-values.

In addition to testing the overall effect of MPC on the full study sample, the study team estimated impacts for several subgroups of interest. The study's analytic approach largely mirrored that used for the full sample findings, but separate impacts were calculated for several subgroups: baseline sexual initiation status, gender, location, and instance of random assignment (to examine potential attenuation effects due to contamination).

2. Implementation study

For the implementation study, the study team used qualitative and quantitative methods to analyze data related to five topics: training and TA received by health educators; youth attendance; implementation fidelity; implementation quality; and youth exposure to TPP and sexual health education and information. The following list summarizes the analytic approach for each topic:

- **Training and TA received by health educators.** The study team synthesized qualitative data from training observations and HTN provider logs, and tabulated responses from the health educator survey.
- **Youth attendance.** Using data entered into the fidelity and attendance logs, the study team calculated the average attendance rate and the percentage of youth attending certain benchmarks for the impact analytic sample overall and by location. In addition to overall attendance rates, the team reports attendance rates for whether youth attended the five key lessons (Lessons 8 and 11 through 14) that the MPC developer and HTN deemed critical for improving behavioral outcomes.
- **Implementation fidelity.** Using data entered by health educators in the fidelity and attendance logs, the study team calculated the percentage of lessons with all activities fully completed, with an addition to the curriculum, and with an adaptation to the curriculum. The team supplemented the quantitative findings with qualitative data from health educators' open-ended responses in the logs.
- **Implementation quality.** The study team calculated the mean scores on several questions from the observations of MPC implementation; these questions were scale measures, with scores ranging from 1 (poor) to 5 (excellent). The team also used qualitative data from youth focus groups to assess quality.
- **Youth exposure to TPP and sexual health education and information.** The study team tabulated information from the student survey and drew on qualitative data from interviews of control group staff.

The team coded qualitative data from interviews and focus groups in NVivo to identify common themes and experiences across respondents.

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

IV. Implementation Findings

The implementation data help explain the extent to which MPC was delivered with fidelity and quality, and the experiences of health educators and youth in the study. They also provide the data necessary to assess the difference between experiences of the MPC and control groups. The subsections that follow describe the training and TA MPC health educators received, youth attendance at MPC lessons, implementation quality and fidelity, and youth exposure to teen pregnancy prevention and sexual health information. Overall, the study found that (1) health educators were well trained and supported: (2) the program was largely implemented as intended; (3) youth attendance was generally high; and (4) there was a clear difference in the amount of sexual health content the MPC and control groups received. Additional details and findings are in Appendix A.

A. Health educator training and technical assistance

As expected, HTN conducted in-person trainings for all 11 health educators who implemented MPC for the study. HTN conducted each training with groups of health educators over two days, except for one training conducted one on one with a single health educator over one day. The study team observed two trainings that included health educators from three cities. Health educators were typically engaged with the material and had time to practice their facilitation skills and receive feedback from trainers and other participants. Eight of the nine health educators who completed the staff survey reported that they felt very prepared to implement MPC after the HTN training.

All 11 health educators also participated in follow-up TA calls with HTN during implementation. In their first year of implementation, health educators participated in two TA calls. During these calls, HTN and the health educators brainstormed solutions to challenges they experienced during implementation, including scheduling and classroom management issues, and potential deviations from the curriculum as written. For example, health educators asked for guidance when implementing some activities with larger classes.

B. Program attendance

Across the three years of the study, attendance rates were generally high, but they varied by city (Exhibit IV.1). On average, students attended 85 percent of the MPC lessons (approximately eight of the 9.5 intended hours of programming).[11] The average attendance rate ranged from 76 percent in St. Louis to 88 percent in Detroit. Most students received the majority of the program; 85 percent of students attended at least 10 of the 14 MPC lessons. However, less than half (46 percent) attended all 14 lessons; this percentage ranged from 61 percent in Mobile to just 20 percent in St. Louis.

Attendance rates for the key MPC lessons also varied considerably by city. More than half of all students (65 percent) attended all five key lessons—Lessons 8 and 11 through 14. The proportion of students who attended these lessons ranged from 47 percent in St. Louis to 73 percent in Mobile. In Cincinnati and Detroit, 60 and 62 percent of students, respectively, attended these lessons.

[11] The sample for the attendance data analysis was limited to individuals who responded to the follow-up survey. Attendance data were missing for 11 respondents.

Exhibit IV.1. MPC attendance was generally high, overall and by city

City	Number of students in treatment group	Percentage of students who attended at least:					Attendance rate (%)
		1 lesson	7 lessons	10 lessons	14 lessons	Lessons 8, 11–14[a]	
Cincinnati	48	100	94	88	42	60	86
Detroit	101	100	99	92	34	62	88
Mobile	505	98	93	88	61	73	87
St. Louis	215	97	87	74	20	47	76
All cities	869	98	93	85	46	65	85

Source: Data that health educators entered into fidelity and attendance logs. An exception is one quarter in Mobile during Cohort 2; data for that quarter came from school attendance records. The results presented here are limited to the sample with attendance data contributing to the impact analyses.

[a] Key lessons that cover the condom demonstration and practice (Lesson 8) and role-playing activities (Lessons 11 through 14).

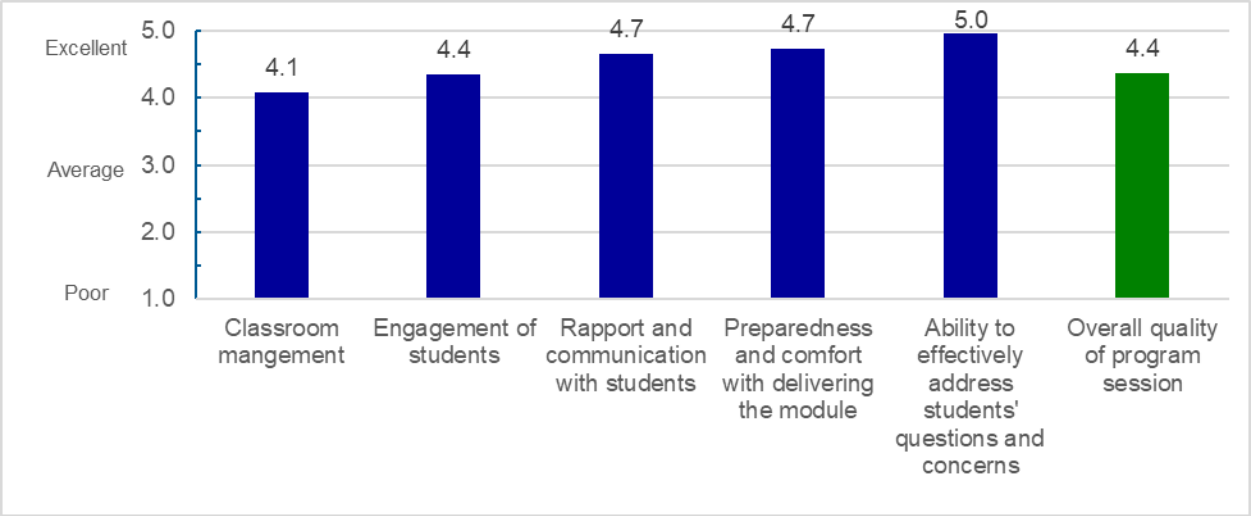
C. Quality

Observation data showed that health educators typically implemented MPC with high quality (Exhibit IV.2). Using a scale from 1 to 5 (with 1 indicating poor performance and 5 indicating excellent), observers rated health educators on various quality metrics.[12] Across all observations, observers rated overall quality as 4.4 out of 5, on average. Other quality metrics included scores for health educator's classroom management, engagement of students, building rapport with students, preparedness and comfort with the curriculum, and ability to address students' questions effectively. The average rating for these other metrics ranged from 4.1 to 5.0. During the observed classes, health educators consistently were prepared and comfortable with program material and able to engage students and address their questions and concerns. These findings were consistent with responses from focus groups, where students generally reported that health educators made the material interesting and easy to discuss, even if some topics were uncomfortable for students.

Although the overall ratings on all quality metrics were high, health educators received the lowest score on the classroom management metric. For most observations, health educators scored a 4 or 5 on this metric (indicating above average performance), but in a quarter of the observations, observers rated health educators with a score of 2 (5 percent of observations) or 3 (20 percent of observations), indicating below average or average performance. According to observers, health educators sometimes experienced challenges with classroom management, noting that youth sometimes held side conversations, were disruptive, or did not pay attention. These observations were consistent with the information that health educators documented in their attendance and fidelity logs, and in interviews with the study team. Regular classroom teachers, who typically stayed in the room while health educators delivered MPC, often assisted with classroom management, and HTN suggested other classroom management strategies during TA calls. However, students in some classrooms still had issues focusing on the lessons throughout implementation.

[12] Observers provided quality scores for each observation; class observations that included two MPC lessons received one set of quality ratings.

Exhibit IV.2. Health educators implemented MPC with high quality in observed classes



Source: Data from observations of 53 MPC lessons (approximately 5 percent of all lessons delivered to youth for the study) conducted by the study team during implementation for Cohorts 2 and 3.

Note: The observation form included descriptive benchmarks for what poor (= 1), average (= 3), and excellent (= 5) scores would look like in the classroom.

The first five bars represent individual quality metrics, whereas the last bar on the far right represents the overall quality score.

D. Fidelity

Overall, health educators implemented MPC with high fidelity. Based on the three metrics the study team examined—completion of activities, additions to MPC content, and adaptations to MPC content—health educators implemented most of the MPC content fully and as intended, without additions or adaptations (Exhibit IV.3).

In 96 percent of implemented lessons, health educators reported that they fully completed all activities. This completion rate ranged from a low of 81 percent in Cincinnati to 98 percent in Mobile. Health educators reported that occasionally they ran out of time to fully complete activities because some activities took longer than expected; class periods were shorter than expected; or classroom management took up a substantial portion of the class period, leaving less time than expected to complete all planned activities. In a few instances, health educators were unable to complete activities as intended because they did not have the technology needed to play the required videos.

Health educators rarely made an addition to the curriculum (3 percent of lessons), although additions were much more common in Cincinnati (40 percent of lessons) compared to other sites (a range of 0.3 to 8 percent of lessons). For instance, in three periods at one school, during the last lesson, the health educator asked students to write down what they had learned and what they would remember from the curriculum. Another health educator used PowerPoint slides during one of the lessons to show definitions of words and phrases used throughout MPC, but HTN staff determined that all of the additions were appropriate and in line with MPC goals.

Exhibit IV.3. Health educators indicated strong fidelity to the MPC curriculum in their fidelity and attendance logs

City	Total number of classrooms where MPC was implemented (across all cohorts)	Percentage of lessons in which the health educator…			Percentage of adaptations that were ….		
		Fully completed all activities	Made an addition	Made an adaptation	Green light	Yellow light	Red light
Cincinnati	3	81	40	33	73	20	7
Detroit	6	86	8	30	52	44	0
Mobile	51	98	0.3	10	92	3	6
St. Louis	15	94	4	13	57	14	29
All cities	**75**	**96**	**3**	**13**	**76**	**14**	**9**

Source: Data that health educators entered into fidelity and attendance logs.

Notes: The study team classified adaptations as green, yellow, or red light, based on input from HTN and ETR's MPC adaptation guidelines. In one case in Detroit, the health educator did not provide enough information in the educator's log to classify the type of adaptation. Percentage of adaptations may not add to 100 percent due to rounding.[13]

Finally, some health educators noted that they made a change to the curriculum in a small proportion of lessons (13 percent). The study team used the MPC adaptation guidelines from ETR, the current distributor of the MPC curriculum, along with feedback from HTN, to classify the adaptations. Most changes were green-light adaptations, which ETR encourages because they are not expected to reduce effectiveness. There were also some yellow- and red-light adaptations, however, which are meant to be implemented carefully or avoided, depending on the adaptation.

- **Green-light adaptations.** Most of the adaptations (76 percent) across all sites were green light, meaning they were minor and did not detract from the content of the curriculum as written. For instance, one health educator did not have youth move their seats to form a circle during activities in the first and last lesson because of the limited space in the classroom. In addition, health educators at one site used their fingers instead of an anatomical penis model to complete the condom demonstration in Lesson 8. The percentage of green-light adaptations varied widely by site, ranging from 52 percent in Detroit to 92 percent in Mobile.
- **Yellow-light adaptations.** Fourteen percent of adaptations across all sites were yellow light, which denote more substantial changes to the curriculum that may make a meaningful difference during implementation. For example, one health educator had to play both parts during the unscripted role-plays in Lesson 13 because the students were uncomfortable performing them. Although this change meant that the health educator skipped part of the activity as written, the educator likely chose the best approach to a trauma-informed practice, particularly as the curriculum advises health educators not to force students to perform role-plays if they do not want to do so. Another health educator was not able to play videos with sound, as intended, in two lessons; in one lesson she played the video without sound and narrated it, whereas in the other she quickly summarized the content verbally. In both lessons, she held the discussion after the video, so youth received the content of the activity despite the technological issues.

[13] OPA currently classifies adaptations as "major" and "minor." More information on this approach is available in the guidance in the Documenting Adaptations Tip Sheet (hhs.gov). ETR continues to use the "green," "yellow," and "red" adaptation categories.

- **Red-light adaptations.** A small proportion (9 percent) of adaptations across all sites were red light, or changes that ETR notes typically should not be implemented. In all cases in this study, the red-light adaptations occurred when a health educator was unable to implement an entire activity (and no alternative related content was added).[14] Most often, health educators made red-light adaptations when they could not play a video because of technical issues, so the students missed the pedagogical element of the video as well as the content it covered. This issue occurred most frequently in St. Louis. As a result, St. Louis had a higher percentage of adaptations categorized as red light (29 percent) compared to the other sites (0 to 7 percent).

E. Youth exposure to teen pregnancy prevention and sexual health information

As intended, the treatment and control groups had different experiences related to TPP programming. Control group youth in Mobile, Detroit, and Cincinnati received business-as-usual classes, most of which did not touch on the topics addressed in MPC as described in interviews with staff. In St. Louis, the control group received the Too Good for Drugs and Violence curriculum, which did not include sexual health content. In Detroit, the control group youth at one of the two schools received some lessons from the Michigan Model, which included a field trip to the health department to highlight birth control options and STI testing.

The information provided in interviews about comparison group programming was corroborated by youth-reported information on the follow-up survey. Youth in the control group reported receiving instruction on fewer TPP and sexual health topics (for example, healthy relationships, birth control methods, and STIs) than youth in the treatment group (effect size [ES] = -0.49, $p < .001$). In addition, when asked about specific types of reproductive health care information they received from a doctor, nurse, or clinic (for example, birth control methods, STIs, and/or the human papillomavirus [HPV] vaccine), youth in the comparison group reported receiving less information (ES = -0.16, $p = .015$) (see Appendix B for more details).

[14] The study team also counted red-light adaptations in calculations for the percentages of activities not fully completed, as they were both adaptations and not fully completed. Following the adaptation guidelines, the team did not consider activities that health educators partially completed as adaptations.

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

V. Impact Findings

This section presents impact results separately for each outcome domain (risk and protective factors, sexual behaviors, and MPC health goals) defined by the MPC logic model in Exhibit II.2. The findings are summarized for each domain, highlighting the statistically significant subset. The section then discusses alignment of the Bayesian posterior probabilities relative to the statistical significance of the observed impacts for the main results. It also summarizes the degree to which findings are robust across alternative data preparation and analytic decisions. The section concludes with a description of the degree to which MPC's effectiveness varied across gender, baseline sexual experience, and subgroups as defined by instance of random assignment in the school (first versus subsequent episode of re-randomization).

Overall, the study found that MPC had (1) several large, favorable, and statistically significant impacts on the outcomes in the risk and protective factor domain; (2) one favorable and statistically significant impact on the outcomes in the sexual behavior domain; and (3) small and nonsignificant impacts for all outcomes in the MPC health goal domain.

A. Risk and protective factors

Across the 10 outcomes examined as risk and protective factors, nine showed statistically significant impacts favoring the MPC program group. The MPC program had favorable and statistically significant impacts on the four measures of knowledge (about HIV/STIs, pregnancy, condoms, and other contraceptives), one measure of beliefs (condoms can be pleasurable), the one measure of attitudes about condoms, and the three measures of self-efficacy (using condoms, negotiating condom use, and refusing sex) (Exhibit V.1).

- **Impacts on knowledge outcomes:** The MPC group had knowledge scores that were significantly higher than the control group on measures of knowledge about HIV (ES = 0.31, $p < .001$), knowledge about pregnancy (ES = 0.17, $p < .001$), knowledge about condoms (ES = 0.55, $p < .001$), and knowledge about other forms of contraception, (ES = 0.11, $p = .037$).
- **Impacts on beliefs:** The MPC group expressed stronger beliefs that condoms can be pleasurable and can be used without ruining the mood or causing embarrassment (ES = 0.15, $p = .012$).
- **Impacts on attitudes about condom use:** The MPC group expressed favorable attitudes about the importance of using condoms every time a person has sex, including their relative ease of use (ES = 0.19, $p < .001$).
- **Impacts on self-efficacy:** The MPC group expressed greater self-efficacy and confidence about using condoms (ES = 0.16, $p = .003$), negotiating the use of condoms (ES = 0.14, $p = .004$), and refusing to have sex (ES = 0.12, $p = .008$).

There was only one outcome in this domain in which there was no statistically significant impact of MPC: the belief that sex may adversely affect future goals. (The control group's score was higher than the MPC group.)

Overall, the findings for the risk and protective factors largely align with expectations. The outcomes measured in this domain are most proximal to the content and activities being taught in MPC. Not surprisingly, nearly all findings in this domain showed large, favorable, and statistically significant impacts.

Exhibit V.1. MPC had many statistically significant and favorable impacts on risk and protective factor domain

Outcome	MPC mean	Control mean	Estimated impact	Effect size (ES)	*p*-value	MPC sample size	Control sample size
Knowledge about HIV	0.51	0.42	0.09***	0.31	< 0.001	844	951
Knowledge about pregnancy	0.58	0.53	0.06***	0.17	< 0.001	856	960
Knowledge about condoms	0.57	0.41	0.16***	0.55	< 0.001	856	959
Knowledge about other forms of contraception	0.18	0.16	0.02*	0.11	0.037	845	952
Belief that sex may adversely affect future goals	0.15	0.18	-0.03	-0.10	0.053	861	967
Belief that condoms can be pleasurable	0.44	0.39	0.05*	0.15	0.012	857	953
Attitudes about condoms	0.70	0.64	0.05***	0.19	< 0.001	870	973
Condom use self-efficacy	0.70	0.64	0.06**	0.16	0.003	854	958
Condom negotiation self-efficacy	0.66	0.61	0.04**	0.14	0.004	856	958
Refusal self-efficacy	0.76	0.72	0.04**	0.12	0.008	863	966

Source: Baseline survey and six-month follow-up survey.

Note: Treatment and control group means are regression adjusted, where higher scores represent better outcomes (for example, more answers correct on a knowledge scale). All regressions adjust for strata fixed effects, a large set of baseline covariates described in Section III, and cluster standard errors at the school*cluster level. All *p*-values are based on a two-sided test. Sample sizes differ across outcomes due to missing outcome data. See Appendix A for descriptions of the individual scales.

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

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

*** Significantly different from zero at the 0.001 level, two-tailed test.

B. Sexual behaviors

Across the eight sexual behavior outcomes examined, the study observed one statistically significant impact. The MPC program group had fewer episodes of sex in the past three months than the control group. The MPC program had no statistically significant impacts on sexual initiation, sex without a condom, or sex without birth control; however, for many outcomes, the direction of the impact suggested a favorable effect of MPC (Exhibit V.2).

- **Impacts on frequency of sex:** On average, youth in the MPC group had sex 2.69 times in the past three months, and the control group had sex 3.99 times (difference = 1.30, ES = -.10, p = .04).

On average, the magnitude of the impacts for the sexual behavior outcomes was smaller than for the risk and protective factors. The average absolute difference in effect size units for sexual behavior outcomes was just 0.05 standard deviations (and a maximum of a 3 percentage point difference). For the outcomes in the risk and protective factor domains, the average absolute difference in effect size units was 0.20 standard deviations—impacts roughly four times the size of those observed for the sexual behavior domain. This smaller magnitude of impacts on the sexual behavior outcomes domain aligns with expectations, in which proximal outcomes were anticipated to have the largest impacts and more distal behavioral outcomes would be smaller, particularly during this short-term follow-up period.

Exhibit V.2. MPC produced one statistically significant impact on sexual behavior, and for many outcomes, the direction suggested a favorable effect

	MPC mean	Control mean	Estimated impact	Effect size	*p*-value	MPC sample size	Control sample size
Ever any sex (vaginal, oral, or anal)	0.50	0.47	0.03	0.06	0.116	827	925
Any sex in past three months	0.29	0.30	-0.01	-0.01	0.775	815	914
Times having any sex in past three months	2.69	3.99	-1.30*	-0.10	0.043	811	908
Count of vaginal sex partners in past three months	0.50	0.68	-0.18	-0.06	0.403	775	868
Any sex without a condom in past three months	0.19	0.20	-0.01	-0.02	0.692	813	901
Times having any sex without a condom in past three months	1.53	2.10	-0.57	-0.05	0.267	813	901
Sex without birth control in past three months	0.11	0.13	-0.02	-0.07	0.178	760	846
Times having sex without birth control in past three months	0.82	0.87	-0.05	-0.01	0.856	760	846

Source: Baseline survey and six-month follow-up survey.

Note: Treatment and control group means are regression adjusted. Impacts on binary outcomes are estimated using the linear probability model, with standard errors adjusted to account for heteroskedasticity. All regressions adjust for strata fixed effects, a large set of baseline covariates described in Section III, and cluster standard errors at the school*cluster level. All *p*-values are based on a two-sided test. Sample sizes differ across outcomes due to missing outcome data. See Appendix A for descriptions of the individual scales.

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

C. MPC health goals

MPC had no statistically significant impact on either of its targeted health outcomes—pregnancy and STI diagnosis (Exhibit V.3). In both the treatment and control groups, about 4 percent of students reported ever having been pregnant at follow-up, and just under 10 percent in each group (8 percent of youth in the treatment group and 9 percent in the control group) reported having had an STI diagnosis in the past three months.

The impacts for the MPC health goal outcomes were the smallest in magnitude relative to the risk and protective factors and sexual behavior outcomes. The average absolute difference for these outcomes was just 0.03 standard deviations (a single percentage point difference for each outcome), again aligning with expectations that these most distal outcomes would be the ones showing the smallest impacts at the short-term follow-up survey.

Exhibit V.3. MPC did not produce any statistically significant impacts on the distal health goals

Outcome	MPC mean	Control mean	Estimated impact	Effect size	*p*-value	MPC sample size	Control sample size
Ever pregnant	0.04	0.04	0.01	0.04	0.384	782	853
Any STI	0.08	0.09	-0.01	-0.03	0.615	822	928

Source: Baseline survey and six-month follow-up survey.

Note: Treatment and control group means are regression adjusted. Impacts on binary outcomes are estimated using the linear probability model, with standard errors adjusted to account for heteroskedasticity. All regressions adjust for strata fixed effects, a large set of baseline covariates described in Section III, and cluster standard errors at the school*cluster level. All *p*-values are based on a two-sided test. Sample sizes differ across outcomes due to missing outcome data. See Appendix A for descriptions of the individual scales.

D. Bayesian interpretation of the impact findings

For all statistically significant outcomes in the risk and protective factor domain, the Bayesian posterior probability strongly suggests MPC had a favorable impact on the outcome. The probability that MPC improved outcomes in this domain was at least 99 percent for each of the impacts that were statistically significant.

The Bayesian posterior probabilities also corroborate the statistically significant favorable impact on the number of times having sex (98 percent probability that MPC had a favorable impact). Furthermore, even though the study found no statistically significant impacts on other sexual behaviors, the Bayesian posterior probabilities suggest MPC may have moved several of these outcomes in the desired direction. These probabilities suggest MPC reduced the likelihood of having sex without birth control (90 percent probability), reduced the number of times youth had sex without a condom (87 percent probability), and reduced the number of vaginal sex partners (80 percent probability). In contrast, the Bayesian posterior probability suggests a high probability (92 percent) that youth in the MPC group had initiated sex at a greater rate than the control group during the follow-up period; however, the magnitude of the finding is small (3 percentage points, relative to a control group prevalence rate of 47 percent).[15] Appendix C presents the Bayesian posterior probabilities for all impact estimates.

E. Results are robust to alternative analysis approaches

As described in Section III, the study team assessed the robustness of the results using several different approaches. The findings were corroborated across multiple approaches. Across six additional analyses using different modeling choices and methods for handling missing or inconsistent data, the impact findings were consistent in impact sign, magnitude, and statistical significance. The large, favorable, and significant impacts on the outcomes in the risk and protective factor domain were replicated almost uniformly across each approach. Furthermore, the statistically significant behavioral finding from the main analysis on the number of times a respondent had sex in the past three months was observed in two other approaches; in the other four analyses, the average *p*-value of the impact for that outcome ranged from 0.06 to 0.12, and the direction of the impact was favorable in all specifications. In addition, the Bayesian posterior probability of a favorable impact on number of times a respondent had sex in the past

[15] As described in the section on recruitment in Appendix B, the study purposively conducted recruitment of schools in areas with high rates of sexual initiation and teen birth rates. The initiation rates observed in this study are markedly higher than the nationwide sexual initiation rates of 19.2 percent for 9th graders and 33.6 percent for 10th graders reported in the 2019 data from the Youth Risk Behavior Survey (Centers for Disease Control and Prevention 2021).

three months was at least 93 percent for all sensitivity analyses, further corroborating the robustness of this finding. A full summary of the results for these analyses is in Appendix C.

F. Subgroup findings

The study team examined the effect of MPC on several key subgroups to understand the extent to which program impacts varied. The team looked at differences by site, gender, instance of random assignment (first or subsequent time the school was randomized to condition), and whether youth were sexually experienced at baseline. The study team was particularly interested in whether the program was effective in changing sexual behavior or MPC health goal outcomes (pregnancy and STI incidence) among sexually experienced youth, given that the original evaluation of MPC showed the program was most effective at improving sexual behavior outcomes among this subgroup. The study's subgroup sample sizes ranged from 313 to 1,150 across outcomes and included roughly 806 individuals, on average, for each contrast tested. Because sample sizes were smaller, the statistical precision of the subgroup impact estimates was limited relative to the full sample analysis. For a full summary of the results for these analyses, see Appendix C.

Across all subgroups examined, there were no statistically significant impacts observed for any sexual behavior or MPC health goal outcomes. Although not statistically significant, the magnitude and direction suggest that the program may have been more effective in improving many sexual behavior and MPC health goal outcomes for sexually experienced youth than inexperienced youth—a finding that aligns with the original Jemmott et al. (1998) study. In particular, the statistically significant reduction in sexual frequency observed in the main sample appears to be largely driven by the sexually experienced subgroup. In addition, though also not significant, the magnitude and direction of impact estimates suggest that the program may have had more favorable results for male students than female students for many sexual behavior and MPC health goal outcomes.

The study team observed several statistically significant impacts on the risk and protective factor outcomes for the subgroups examined. For both sexually experienced and inexperienced youth, MPC caused significant improvement in knowledge outcomes; for the sexually inexperienced sample, the MPC program produced significantly better scores with respect to condom beliefs, attitudes about condoms, and condom self-efficacy. Most of the full sample findings on risk and protective factors remained statistically significant when examined across the male and female subgroups. The magnitude of the impacts tended to be larger and more frequently statistically significant in the second (or third) instance of random assignment relative to the first instance for outcomes in the risk and protective factor domain. This result may have been because health educators were more comfortable after their first instance of implementation. Finally, the team observed several of the favorable impacts on risk and protective factors in the collection of 18 clusters in Mobile, although none of the impacts on these variables was statistically significant when examined among the seven clusters for the St. Louis schools (possibly a result of low power). The team did not attempt to examine the effectiveness of MPC in Detroit or Cincinnati, given the small number of clusters in those locations.

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

VI. Conclusion

This study provided a rigorous evaluation of the effectiveness and implementation of MPC. The logic model for the MPC program suggests that it first affects risk and protective factors for risky sexual behavior, including knowledge, attitudes, beliefs, and skills and self-efficacy. Improvement in these proximal outcomes subsequently is expected to influence sexual behaviors, and eventually the health goals of the MPC program—reducing teen pregnancy and STIs. Overall, in this study, health educators implemented MPC with high quality and fidelity, and lessons were well attended. There was a strong contrast observed in the experiences between the treatment and control youth.

About six months after the end of the program, the study showed several large, statistically significant, and favorable impacts on nearly all of the risk and protective factors examined in this evaluation. The MPC program significantly improved knowledge of HIV/STIs, pregnancy, condoms, and other contraceptives; attitudes and beliefs about condoms; and self-efficacy in using condoms, negotiating condom use, and refusing sex.

Among the 10 sexual behavior and MPC health goals examined (pregnancy and STI prevention), the study observed one statistically significant impact. The MPC group reported significantly fewer episodes of sex in the past three months relative to the control group. The magnitude of the impact estimates for other behavioral and MPC health goal outcomes were small and nonsignificant, but the direction of most findings suggested a favorable effect of MPC.

The finding that MPC produced larger and more significant impacts on the risk and protective factors than the more distal behavioral outcomes is consistent with the logic model for the program. The proximal outcomes aligned well with the content and activities provided by MPC and were expected to change the most by the time of the short-term, six-month follow-up assessment. Although a long-term follow-up assessment originally was planned when the study was designed, the contract option was not exercised, so all outcomes were examined at the short-term follow-up period. This departure from the design set the stage for the findings shown here: favorable and significant impacts on risk and protective factors that may have only begun to manifest themselves into significant impacts on behavior.

Across all subgroups examined, there were no statistically significant impacts observed for any sexual behavior or MPC health goal outcomes. The magnitude and direction of the impacts suggest the program may have been more effective in improving several sexual behavior and MPC health goal outcomes for sexually experienced youth—a finding that aligns with the original Jemmott et al. (1998) study. In addition, the magnitude and direction of the effects suggest the program may have had more favorable results for male students than female students for many sexual behavior and MPC health goal outcomes. However, these estimates of program impacts for subgroups were less precise than the full sample findings, based on the smaller sample sizes.

Although the current study did observe one favorable and statistically significant impact of MPC on a behavioral outcome, it did not fully replicate the findings from the original study—specifically, that MPC was effective at reducing a range of several risky sexual behavior outcomes in the short term after the end of the intervention. There were several notable differences between the original Jemmott et al. (1998) evaluation and the current one, which may offer some insight into why the original study may have shown more favorable impacts on behavioral outcomes. The original evaluation occurred during the height of the HIV epidemic, when sexual health prevention programming was far rarer and teen birth rates were substantially higher than today, so there could have been more opportunity for changes in youth behavior at that time. There also could have been differences in responsiveness to programming among youth, given that the first study involved youth who volunteered to participate in programming on weekends outside of school versus high school students who participated during a required school class. Although there was a strong contrast in the experiences between the treatment and control youth tested in this study, it may not have been as strong as in the original study, in which youth in the control group did not receive any sexual health programming and the program developer oversaw program delivery. In

combination, these differences potentially could have reduced the opportunity for the current study to produce large impacts on behavioral outcomes. Looking across the two studies, these differences may explain the more limited program impacts this study observed on behavioral outcomes.

Overall, the MPC evaluation found favorable program effects, including the type of findings that TPP evidence review criteria can use to characterize MPC as a program with evidence of effectiveness. The program showed consistently favorable, large, and statistically significant impacts on the outcomes identified in the logic model as proximal short-term outcomes. Importantly, for behavioral outcomes, it showed one significant favorable impact, several nonsignificant impacts, and no significant unfavorable impacts. Ideally, the study would have produced long-term follow-up data to understand whether the impacts on risk and protective factors subsequently produced significant favorable changes on more of the sexual behaviors and, eventually, on MPC health goals as hypothesized by the logic model. Thus, more research into the longer-term effectiveness of MPC may be needed to address this remaining gap in the literature. However, even without these longer-term findings, the current study provides evidence to suggest that as implemented today, MPC should continue to be considered an evidence-based program that favorably affects behavioral outcomes.

VII. References

Abt Associates. “Teen Pregnancy Prevention (TPP) Replication Study: Impact Study Design Report.” Cambridge, MA: Abt Associates, 2015.

Centers for Disease Control and Prevention (CDC). “HIV/AIDS Surveillance Report: US HIV and AIDS Cases Reported Through June 1996.” Atlanta, GA: Department of Health and Human Services, Public Health Services, Center for Infectious Disease, Centers for Disease Control and Prevention, 1996.

Centers for Disease Control and Prevention (CDC). “1991−2019 High School Youth Risk Behavior Survey Data.” Available at http://yrbs-explorer.services.cdc.gov/. Accessed September 28, 2021.

Coyle, K.K., J.R. Glassman, H.M. Franks, S.M. Campe, J. Denner, and G.M. Lepore. “Interventions to Reduce Sexual Risk Behaviors Among Youth in Alternative Schools: A Randomized Controlled Trial.” *Journal of Adolescent Health*, vol. 53, no. 1, 2013, pp. 68–78. doi:10.1016/j.jadohealth.2012.12.012. PMID: 23562220; PMCID: PMC3691297.

Deke, J., and M. Finucane. “Moving Beyond Statistical Significance: The BASIE (BAyeSian Interpretation of Estimates) Framework for Interpreting Findings from Impact Evaluations.” OPRE Report #2019-35. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2019.

ETR Associates. “Making Proud Choices! Fifth Edition: Adaptations Guidelines.” Scotts Valley, CA: ETR, 2017.

Farb, Amy, and Amy L. Margolis. “The Teen Pregnancy Prevention Program (2010−2015): Synthesis of Impact Findings.” *American Journal of Public Health*, vol. 106, no., 2016, pp. S9−S15. doi.org/10.2105/AJPH.2016.303367.

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian data analysis (3rd ed.)*. CRC Press, 2013.

Goesling, Brian, Silvie Colman, Christopher Trenholm, Mary Terzian, and Kristin Moore. “Programs to Reduce Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors: A Systematic Review.” *Journal of Adolescent Health*, vol. 54, no. 5, 2014, pp. 499−507.

Goesling, Brian. “Making Sense of Replication Studies: Guidance for Teen Pregnancy Prevention Researchers.” Assistant Secretary for Planning and Evaluation (ASPE) Research Brief, May 2015.

Goesling, Brian, Joanne Lee, Robert G. Wood, and Susan Zief. “Adapting an Evidence-Based Curriculum in a Rural Setting: The Longer-Term Impacts of Reducing the Risk in Kentucky.” OPRE Report #2018-27. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018.

Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. “Statistical Tests, *p*-Values, Confidence Intervals, and Power: A Guide to Misinterpretations.” *European Journal of Epidemiology*, vol. 31, no. 4, 2016, pp. 337–350.

Haskins, R., and G. Margolis. *Show Me the Evidence: Obama’s Fight for Rigor and Evidence in Social Policy*. Washington, DC: Brookings Institution Press, 2015.

Herrmann, M., M. Clark, S. James-Burdumy, C. Tuttle, T. Kautz, V. Knechtel, D. Dotter, C.S. Wulsin, and J. Deke. “The Effects of a Principal Professional Development Program Focused on Instructional Leadership.” NCEE 2020-0002. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2019.

Jemmott, J.B., L.S. Jemmott, and G.T. Fong. “Abstinence and Safer Sex HIV Risk-Reduction Interventions for African American Adolescents: A Randomized Controlled Trial.” *Journal of the American Medical Association*, vol. 279, no. 19, 1998, pp. 1529–1536.

Jemmott, J. B., L.S. Jemmott, G.T. Fong, and K.H. Morales. “Effectiveness of an HIV/STD Risk-Reduction Intervention for Adolescents When Implemented by Community-Based Organizations: A Cluster-Randomized Controlled Trial.” *American Journal of Public Health*, vol. 100, no. 4, 2010, pp. 720–726.

Juras, R., E. Tanner-Smith, M. Kelsey, M. Lipsey, and J. Layzer. “Adolescent Pregnancy Prevention: Meta-Analysis of Federally Funded Program Evaluations.” *American Journal of Public Health*, vol. 109, no. 4, 2019, pp. e1–e8.

Kappeler, Evelyn M., and Amy Feldman Farb. “Historical Context for the Creation of the Office of Adolescent Health and the Teen Pregnancy Prevention Program.” *Journal of Adolescent Health*, vol. 54, no. 3, Supplement, 2014, pp. S3–S9.

Kaye, K., K. Suellentrop, and C. Sloup. “The Fog Zone: How Misperceptions, Magical Thinking, and Ambivalence Put Young Adults at Risk for Unplanned Pregnancy.” Washington, DC: Power to Decide (formerly The National Campaign to Prevent Teen and Unplanned Pregnancy), 2009.

Kautz, Tim, and Russell Cole. “Selecting Benchmark and Sensitivity Analyses.” Evaluation Technical Assistance Brief. Submitted to the Office of Adolescent Health. Princeton, NJ: Mathematica Policy Research, September 2017.

Lewandowski, D., D. Kurowicka, and H. Joe. “Generating random correlation matrices based on vines and extended onion method.” *Journal of Multivariate Analysis,* vol. 100, 2009, pp. 1989–2001.

Markham, C.M., S.R. Tortolero, M.F. Peskin, R. Shegog, M. Thiel, E.R. Baumler, R.C. Addy, S.L. Escobar-Chaves, B. Reininger, and L. Robin. “Sexual Risk Avoidance and Sexual Risk Reduction Interventions for Middle School Youth: A Randomized Controlled Trial.” *The Journal of Adolescent Health*, vol. 50, no. 3, 2012, pp. 279–288. Available at https://doi.org/10.1016/j.jadohealth.2011.07.010. Accessed October 8, 2021.

Markham, C.M., M.F. Peskin, R. Shegog, E.R. Baumler, R.C. Addy, M. Thiel, S.L. Escobar-Chaves, L. Robin, and S.R. Tortolero. “Behavioral and Psychosocial Effects of Two Middle School Sexual Health Education Programs at Tenth-Grade Follow-Up.” *The Journal of Adolescent Health*, vol. 54, no. 2, 2014, pp. 151–159.

Mathematica Policy Research. “Identifying Programs that Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors: Review Protocol Version 5.0.” 2014. Available at https://tppevidencereview.youth.gov/ReviewProtocol.aspx. Accessed October 8, 2021.

Murphy, Lauren, Lara Hulsey, and Susan Zief. “PREP Performance Measures Final Report.” OPRE Report #2021-12, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2021.

Patton, M. Q. “Qualitative Research and Evaluation Methods (3rd ed.).” Thousand Oaks, CA: SAGE Publications Inc., 2002.

Pound, P., R. Langford, and R. Campbell. “What Do Young People Think About Their School-Based Sex and Relationship Education? A Qualitative Synthesis of Young People’s Views and Experiences.” *BMJ Open*, vol. 6, no. 9, 2016. doi:10.1136/bmjopen-2016-011329.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. “What to Do When Data Are Missing in Group Randomized Controlled Trials.” NCEE 2009-0049. Washington, DC: National Center for Education Evaluation and Regional Assistance, 2009.Ritchie, J., and L. Spencer. “Qualitative Data Analysis for Applied Policy Research.” In *The Qualitative Researcher's Companion,* edited by A. M. Huberman and M. B. Miles (pp. 305–329). SAGE Publications, Inc., 2002.

Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc., 1987.

Valentine, Jeffrey C., Anthony Biglan, Robert F. Boruch, Felipe González Castro, Linda M. Collins, Brian R. Flay, Sheppard Kellam, Eve K. Mościcki, and Steven P. Schinke. “Replication in Prevention Science.” *Prevention Science*, vol. 12, no. 2, 2011, pp. 103–117.

Wasserstein, Ronald L., and Nicole A. Lazar. “The ASA’s Statement on *p*-Values: Context, Process, and Purpose.” *American Statistician*, vol. 70, no. 2, 2016, pp. 129–133.

What Works Clearinghouse. “Standards Handbook, Version 4.1.” Washington, DC: Institute of Education Sciences, 2020.

Zief, Susan, Rachel Shapiro, and Debra Strong. “The Personal Responsibility Education Program (PREP): Launching a Nationwide Adolescent Pregnancy Prevention Effort.” OPRE Report #2013-37. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, October 2013.

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

Appendix A: Details on Data, Methods, and Analyses

This appendix provides additional details for the MPC impact and implementation study design, data sources, and methods. It first describes features of the overall design and the impact study and concludes with a section on implementation data sources and methods.

A. Description of re-randomization feasibility

This evaluation used a cluster RCT design to assess the effectiveness of MPC, with high schools within cities randomly assigned to one of two conditions: (1) MPC implemented by health educators or (2) business as usual. Before random assignment, each school was able to select the classroom setting for the study where MPC would be delivered if the school were randomized to receive it or where business-as-usual programming would occur. Schools most frequently selected health class as the target classroom for the study across the three cohorts.[16]

Health educators from local partner organizations received training and TA from HTN and implemented the program.[17] The study was implemented over the course of three school years, with youth in each year considered a separate evaluation cohort. Several of the schools participated in the study across multiple cohorts. These schools were re-randomized to condition each year they participated. As a result, many schools switched conditions across cohorts. This re-randomization of schools to condition each year was feasible because (1) different students were eligible for participation in the evaluation each year and (2) the health educators could move across schools between cohorts (that is, the treatment was impermanent within a school).

1. Different student populations in each cohort: The classes on which the evaluation focused were typically either a one-semester or one-quarter class, so the eligible students within a school were unique each year. Although a small number of students may have repeated the course in multiple years, they were considered eligible for the study only the first time they were enrolled in the target class.
2. The program could "move" with the health educators: The MPC program was a localized, impermanent program, expected to affect students only during the semester or quarter when it was implemented and they had an opportunity to participate. This program does not seek to change the school environment as a whole, making it possible to remove the program from a school at the end of each cohort.

The section below on the approach for estimating impacts describes how the study re-randomized schools to condition in the benchmark and sensitivity analyses.

B. Recruitment

Beginning in spring 2016, the study team worked with OPA to identify partner organizations to participate in the study. A variety of candidate provider organizations that had experience providing MPC in local school districts were considered. The team prioritized organizations working in urban areas where the teen birth or STI rates were markedly higher than state or national averages to ensure the study was serving a high-risk population.

[16] As a result of this design feature, the services available or provided to the comparison group were different at each school. However, the comparison programming never included a competing evidence-based program intended to affect sexual behavior outcomes.

[17] The partner organizations were the Mobile County Health Department, the YMCA of Metropolitan Detroit, Planned Parenthood of Southwest Ohio in Cincinnati, and Better Family Life in St. Louis. For the first cohort in Detroit (during the 2017–2018 school year), the health educator was an independent contractor.

The team ultimately recruited implementation partner organizations, districts, and schools in four cities or metropolitan areas: (1) Mobile, Alabama; (2) Detroit, Michigan; (3) Cincinnati, Ohio; and (4) St. Louis, Missouri. Exhibit A.1 presents the MPC study sites and partners.

Exhibit A.1. MPC study sites and partners

City and state	Implementation partner organization	Total number of health educators who implemented MPC	Number of Schools
Mobile, Alabama	Mobile County Health Department	3	6
Detroit, Michigan	YMCA of Metropolitan Detroit	2	2
Cincinnati, Ohio	Planned Parenthood of Southwest Ohio	1	1
St. Louis, Missouri	Better Family Life	5	6

Mobile, Alabama. Beginning in September 2016, the study team worked with Mobile County Health Department, its implementation partner organization, to recruit the Mobile County Public School System. The school system identified six high schools to prioritize for recruitment and participation in the study. All six schools agreed to participate in the study during the last quarter of the 2016–2017 school year (Cohort 1). All six schools agreed to participate in the study during the 2017–2018 (Cohort 2) and 2018–2019 (Cohort 3) school years, as well. Due to the quarter system in Mobile and the desire to minimize interruptions at the schools while still holding constant data collection activities on a common timeline across conditions, consent and baseline data collection for Cohorts 2 and 3 were conducted at the start of the school year for all students in 10th grade who were likely to take health in Quarters 2, 3, or 4, or Reserve Officers' Training Corps (ROTC) in Semester 1 (see below). At the start of each quarter or semester, a final class roster was obtained to determine the eligible sample. Schools were randomized to condition separately within each cohort of participation.

In five of the six schools in Mobile, the evaluation focused on students in health class (primarily 10th graders). In one school, the evaluation focused on students in ROTC (primarily 9th graders). In all instances, MPC was provided by health educators from the Mobile County Health Department. When assigned to the control condition, schools continued with their regular health or ROTC programming. The business-as-usual health programming provided through health class was expected to include basic sexual health content, but this information was covered in fewer classes and did not include much of the more practical, experiential content offered through MPC. Sexual health content was not provided during business-as-usual ROTC programming.

Detroit metropolitan area, Michigan. Beginning in June 2017, the study team worked with YMCA of Metropolitan Detroit, its implementation partner organization, to identify potential districts and schools in and around Detroit. The YMCA of Metropolitan Detroit identified two high schools in two different districts in the greater metropolitan area; both agreed to participate in the study during the 2017–2018 (Cohort 2) and 2018–2019 (Cohort 3) school years. Consent and baseline data collection for these schools occurred during the fall semester of each school year.

The two schools were randomized to condition so that one school received MPC during Cohort 2 and the other served as the control group; their conditions then swapped for Cohort 3.

In one of the Detroit schools, the evaluation focused on students in health class (primarily 9th and 10th graders), and a health educator from the YMCA of Metropolitan Detroit provided MPC to them. When assigned to the control condition, the school continued with its regular health programming. During typical health classes, teachers at this school used the Michigan Model, a curriculum developed in Michigan for K–12 public schools, to meet the State Sexual Health Standards. For high school youth, the Michigan Model includes the Healthy and Responsible Relationships curriculum, which can include up to 22 lessons covering healthy relationships and sexual health, including condoms and contraception. Some

lessons are short, so multiple lessons can fit within a single class period, and teachers have some discretion on which lessons they implement. Although the Michigan Model covers some of the same information as MPC, including STIs/HIV and pregnancy prevention, it provides a limited amount of information on many different topics and focuses more broadly on healthy relationships. When this school was assigned to the treatment condition, lessons on STIs/HIV and pregnancy prevention were removed from the curriculum for the year.

In the other Detroit school, the evaluation focused on students in Leadership Education 1 (a grade 9 ROTC class), and a health educator working independently (and previously a trainer for HTN) provided MPC to them. When assigned to the control condition, the school continued with its regular ROTC programming. In ROTC, the students had very little additional health programming aligned with MPC. Students may have received some sexual health education if they were enrolled in a physical education class at the same time as ROTC. Although students could meet their physical education requirements by taking the ROTC class, they could also enroll in a physical education class, which included a section on health. During the section on health, teachers used the version of the Michigan Model that covers abstinence and condoms for disease risk reduction.

Cincinnati, Ohio. Beginning in June 2017, the study team worked with Planned Parenthood of Southwest Ohio, its implementation partner organization, to identify schools to participate in the study. One school agreed to participate during the 2017–2018 school year and again in the 2018–2019 school year. The school understood that it would receive MPC in one school year and serve as a control school in the other year of the study, and that the year in which it would receive MPC would be determined randomly. Consent and baseline data collection occurred in the spring semester of each year.

In this school, the evaluation focused on youth in health class (primarily 9th graders); a Planned Parenthood health educator provided the programming. When assigned to the control condition, the school continued with its regular health programming, which included basic sexual health content, but this information was covered in fewer lessons and without the more practical, experiential content offered by MPC. During the years in which the school participated in the study, a one-week version of Choosing the Best, an abstinence curriculum that focuses on healthy relationships, was also offered to youth in health classes. Importantly, because this program was experienced by youth in both treatment and control conditions, it did not contribute to the estimate of the effect of MPC (though it did create the potential for a more saturated environment where treatment versus control experiences might be attenuated).

St. Louis, Missouri-Illinois metropolitan area. Beginning in spring 2018, the study team worked with Better Family Life (BFL), its implementation partner organization, to identify potential schools in the St. Louis area. These schools could have included those in the St. Louis Public Schools district as well as other surrounding districts in the St. Louis metropolitan area in both Missouri and Illinois.

Initially, one school in St. Louis Public Schools agreed to participate in the study in the spring 2018 semester and agreed again for the 2018–2019 school year. Like the high school in Cincinnati, this school understood that it would receive MPC in one school year and serve as a control school in the other year of the study, and that the year in which it would receive MPC would be determined randomly. Five additional schools were recruited between June and December 2018 for participation in the evaluation during the 2018–2019 school year (Cohort 3). The consent and baseline data collection schedule varied depending on when the schools were recruited. In four schools, the consent and baseline collection occurred in the fall 2018 semester; in two schools, it occurred in the spring 2019 semester.

In St. Louis, the evaluation focused on several different classes, depending on the school. At three of the high schools, it focused on students in 9th and 10th grade gym class; health educators from BFL provided the programming. In one high school, the evaluation focused on 9th and 10th grade students in writing, theater, and computer skills classes; health educators provided the programming. In the remaining two high schools, the evaluation focused on students in 9th grade history or career development, respectively; health educators provided programming to them.

When assigned to the control condition, schools in St. Louis received Too Good for Drugs and Violence, a different curriculum that did not cover sexual health content. Trained health educators from BFL delivered this program as well.

C. Random assignment results

Exhibit A.2 provides the random assignment status of all schools included in the evaluation.

Exhibit A.2. Random assignment results (n = 31 clusters assigned to condition)

School	City	Treatment status, Cohort 1	Treatment status, Cohort 2	Treatment status, Cohort 3	Classes selected	Primary grade of selected class
1	Mobile	Treatment	Control	Treatment	Health	10
2	Mobile	Treatment	Control	Control	Health	10
3	Mobile	Treatment	Treatment	Treatment	Health	10
4[a]	Mobile	Control	Control	Control	ROTC	9
5	Mobile	Control	Treatment	Control	Health	10
6	Mobile	Control	Treatment	Treatment	Health	10
1[b]	Cincinnati	n.a.	Treatment	Control	Health	9
1[c]	Detroit	n.a.	Control	Treatment	Health	9–10
2[c]	Detroit	n.a.	Treatment	Control	Leadership education	9
1[b,d]	St. Louis	n.a.	Control	Treatment	Physical education	9–10
2	St. Louis	n.a.	n.a.	Control	Writing; theater; computer skills	9–10
3	St. Louis	n.a.	n.a.	Treatment	Physical education	9–10
4	St. Louis	n.a.	n.a.	Control	History	9
5	St. Louis	n.a.	n.a.	Treatment	Career development	9
6	St. Louis	n.a.	n.a.	Control	Physical education	9–10

[a] This school did not switch conditions (it was assigned to the control group in each of the three instances of random assignment), which may have produced two unexpected baseline differences that resulted from random sampling error. First, because it was the only school in Mobile that targeted 9th graders, the control group had slightly more 9th graders and fewer 10th graders. Second, this school had baseline data collection in fall of the school year and follow-up data collection in the spring, whereas the other schools had some students with baseline and follow-up data collection occurring in the fall of adjacent years. This data collection timing produced a longer follow-up period for the treatment group, on average. The study team conducted a sensitivity analysis (not shown in Appendix C) in which it excluded data from this school; the direction and significance of the findings for all analyses were identical to the benchmark approach. Additional tests of baseline equivalence (not shown) that excluded this school showed significant reductions in the two differences described above, so they fell below typical sensitivity thresholds.

[b] Each of these schools served as its own stratum, with varying treatment statuses across Cohorts 2 and 3.

[c] These schools served as their own strata, with varying treatment statuses across Cohorts 2 and 3, and randomization was constrained so that only one school would receive MPC during each cohort.

[d] For this school, baseline surveys were administered in the spring for Cohort 2 but in the fall for Cohort 3, consistent with baseline survey timing for other schools in St. Louis. This change in survey timing also contributed to the unexpected difference in time between baseline and follow-up noted above in table note a.

n.a. = not applicable (that is, the school was not enrolled in the study during that cohort); ROTC = Reserve Officers' Training Corps.

D. Youth consent process details

Several weeks before the start of programming, the study team worked with staff in all treatment and control schools to collect consent forms.[18] A letter of support for the study from the district was included with each consent form sent home to students. Parents were given the option to return the hard copy form or log onto a website and consent electronically. The consent process was conducted until at least 90 percent of parents returned consent forms (regardless of whether they actually consented to have their child participate in the study). Students received a small gift bag (valued at approximately $5) with items such as ear buds, pencils, and nylon backpacks for returning a form. The consent process was the same across treatment and control schools, and was not associated with the future offer of MPC in the treatment schools; parents and students were not aware of a school's treatment status at the time of study consent.

The study team relied on several strategies to reach the goal of receiving 90 percent of consent forms. Study staff established strong relationships with a point person at the school, typically in the front office. They coordinated at least three visits during the consent process to collect completed forms and drop off extra forms when necessary. Classroom teachers encouraged completion, reminding students of the importance of returning the forms. The study team also worked with schools to use systems already in place to reach parents and remind them to complete forms—for example, collecting forms during back-to-school events and using the school's automated text messaging system for reminders.

In some schools, study staff called parents and obtained verbal consent. Study staff made these calls from the school after student dismissal and in the early evening hours, when possible, because parents were more likely to be available and answer a familiar number.

Exhibit A.3 provides the total number of eligible students in each city, by cohort, and the percentage of those eligible who consented to participate in the study.

[18] Consent for study participation occurred after schools were randomly assigned to condition.

Exhibit A.3. Study consent rates, by cohort and city

	Total eligible		Consented	
	n		n (%)	
Cohort/city	Treatment	Control	Treatment	Control
Cohort 1 (spring 2017 implementation)				
Mobile	108	238	90 (83.3%)	184 (77.3%)
Cohort 2 (2017–2018 school year implementation)				
Mobile	361	304	252 (69.8%)	247 (81.3%)
Detroit[a]	62	120	55 (88.7%)	89 (74.2%)
Cincinnati[b]	70		57 (81.4%)	
St. Louis[c]		80		38 (47.5%)
Cohort 3 (2018–2019 school year implementation)				
Mobile	306	380	221 (72.2%)	275 (72.4%)
Detroit[a]	75	54	56 (74.7%)	49 (90.7%)
Cincinnati[b]		52		41 (78.8%)
St. Louis[c]	334	266	279 (83.5%)	205 (77.1%)
Total	**1,316**	**1,494**	**1,010 (76.7%)**	**1,128 (75.5%)**

[a] In Detroit, the two participating schools served as their own strata, with varying treatment statuses across Cohorts 2 and 3; randomization was constrained so that only one school would receive MPC during each cohort.

[b] The one school in Cincinnati served as its own stratum, with varying treatment statuses across Cohorts 2 and 3.

[c] In St. Louis, one school served as its own stratum, with varying treatment statuses across Cohorts 2 and 3.

E. Survey data collection process

Youth in the study completed two waves of self-report surveys: (1) a baseline survey administered before programming began and (2) a follow-up outcome survey approximately six months after the end of programming (approximately nine months after baseline).

The primary administration approach for the surveys was in school and was made by trained data collection staff. Youth completed the survey online using study-provided, Internet-connected smartphones. For follow-up data collection, when youth were not in school on the days of in-person data collection, data collectors attempted to follow up with them outside of school. Youth received invitations by mail and email with information on how to complete the survey online. This contact also provided them with an option to call Mathematica's survey operations center to complete the survey by telephone. Approximately one week after the invitations were sent, telephone interviewers began calling the remaining nonresponders to try to complete the survey by telephone. During the rest of the data collection period, nonresponders received emails and text reminders weekly. To encourage participation, youth who completed the follow-up survey in school were given a $5 gift card for a local area restaurant; youth who completed the survey outside of school were given a $10 Visa gift card.

Among the consented study youth, 81 percent completed the follow-up survey in school, 4 percent completed the survey outside of school by telephone, and 2 percent completed the web survey out of school. Thirteen percent of youth never completed the follow-up survey. These response rates were similar across treatment and control groups for both the baseline and follow-up surveys. Detailed survey response rates by wave, cohort, and city are in Exhibit A.4.

Exhibit A.4. Baseline and follow-up survey response rates, by cohort and city, relative to the number of youth with study consent

	Baseline survey response rate N (%)		Follow-up survey response rate N (%)	
Cohort/city	**Treatment**	**Control**	**Treatment**	**Control**
Cohort 1 (spring 2017 implementation)				
Mobile	82 (91%)	178 (97%)	79 (88%)	167 (91%)
Cohort 2 (2017–2018 school year implementation)				
Mobile	235 (93%)	241 (98%)	232 (92%)	222 (90%)
Detroit[a]	53 (96%)	87 (98%)	53 (96%)	80 (90%)
Cincinnati[b]	55 (96%)		48 (84%)	
St. Louis[c]		34 (89%)		29 (76%)
Cohort 3 (2018–2019 school year implementation)				
Mobile	214 (97%)	258 (94%)	202 (91%)	254 (92%)
Detroit[a]	53 (95%)	49 (100%)	48 (86%)	45 (92%)
Cincinnati[b]		40 (98%)		32 (78%)
St. Louis[c]	258 (92%)	198 (97%)	218 (78%)	171 (83%)
Total	**950 (94%)**	**1,085 (96%)**	**880 (87%)**	**1,000 (89%)**

Note: Response rates reported in parentheses are relative to youth consenting to the study.

[a] In Detroit, the two participating schools served as their own strata, with varying treatment statuses across Cohorts 2 and 3; randomization was constrained so that only one school would receive MPC during each cohort.

[b] The school in Cincinnati served as its own stratum, with varying treatment statuses across Cohorts 2 and 3.

[c] In St. Louis, one school served as its own stratum, with varying treatment statuses across Cohorts 2 and 3.

The study team assessed the threat of bias stemming from follow-up survey nonresponse relative to the initially eligible sample. Assessing sample attrition relative to the initially assigned sample instead of the consented sample is a more conservative analysis and is typically conducted as part of an evidence review. According to WWC attrition standards using the *cautious* boundary, the study had *low* rates of attrition at the follow-up assessment (33.5 percent overall attrition and 0.2 percent differential attrition).

The study team also examined the threat of item nonresponse as a potential source of internal validity bias, given that some youth might skip sensitive items within the survey. There was little item nonresponse among the survey responders; again, there was a *low* threat of attrition associated with all three domains of outcomes (Exhibit A.5).

Exhibit A.5. Survey and outcome response rates

	Treatment	Control	Total (N or response rate)	Differential response rate
Initially assigned sample	1,316	1,494	2,810	
Follow-up survey respondents	880	1,000	1,880	
Survey response rate relative to initially assigned sample	66.6%	66.4%	66.5%	0.2 PP
Responded to risks and protective factor outcomes	876	991	1,867	
Risk and protective factor outcome response rate relative to initially assigned sample	66.6%	66.3%	66.4%	0.2 PP
Responded to sexual behavior outcomes	835	934	1,769	
Sexual behavior outcome response rate relative to initially assigned sample	63.4%	62.5%	63.0%	0.9 PP
Responded to MPC health goal outcomes	846	954	1,800	
MPC health goal outcome response rate relative to initially assigned sample	64.3%	63.9%	64.1%	0.4 PP

PP = percentage point.

F. Survey data collection timing

The study team scheduled survey data collection events so that the timing of baseline data collection was consistent within a location and there was approximately six months between the end of programming and the follow-up period. The timing of MPC and control programming depended on each school's academic calendar and the class targeted for the study (for example, 10th grade health class), with students attending the target classes for a quarter, semester, or year. Because the timing of programming varied by location and school, the timing of follow-up data collection also varied. Students in a cohort received baseline surveys simultaneously—usually in the fall—but received their follow-up surveys roughly six months after the end of programming, which led to variation in the timing of follow-up data collection. Students enrolled in the target class in a fall semester took their baseline survey in the fall and their follow-up survey in the spring (in the same school year). Students rostered to the target class in the spring took their baseline survey in the fall and their follow-up survey the following fall (in the subsequent school year), roughly a year after baseline and with a summer break in between.

The data collection schedule was developed for each school site to accommodate when treatment and control coursework would be offered before random assignment occurred. The study team expected that through random assignment, the length of time between baseline and follow-up would be balanced across all participating treatment and control schools. However, despite the random assignment of schools to condition, an imbalance in length of time between baseline and follow-up emerged across conditions. As noted in Exhibit A.2, the imbalance in timing was due in part to one school in Mobile that did not change conditions during the study, as well as a school in St. Louis that had a different data collection schedule between Cohorts 2 and 3.

Because a difference in the length of time between baseline and follow-up across conditions could plausibly influence survey responses (for instance, sexual initiation or the number of sexual encounters in the past three months), the study's benchmark analysis included regression controls both for follow-up interval length and having a summer break in the interval. The study team also conducted a robustness check that excluded these survey timing variables and other regression controls and found that the results were similar in direction and significance to the benchmark model, indicating that differences in survey timing did not substantively affect the impact findings.

G. Preparation of data for analysis

To simplify the analysis and the interpretation of results, the study team took a number of steps to clean, simplify, and harmonize students' raw survey responses, both within each survey and across the two surveys.

Because many of the analyses accounted for student characteristics such as race, gender, and grade level, the study team sought to avoid missing data for these characteristics wherever possible. In each case, the student's baseline survey response was used when it was available. If the student did not fill out a baseline survey or skipped the response in question, the study team filled in missing values for these data using logical imputation based on other data sources that addressed the same topic. For instance, if students did not indicate their gender at baseline, the team referred instead to their follow-up survey response. If this response also was missing, the team referred to their parental consent form, and finally to school roster data when available. Exhibit A.6 indicates the study's process of logical imputation for different key variables.

Exhibit A.6. Alternative data sources for logically imputing missing student baseline characteristics

Student characteristics	Alternative data sources (when missing from baseline survey)
Gender	Student follow-up survey, study participation parental consent form, school roster
Date of birth	Follow-up survey response, study participation parental consent form, school roster
Grade level at baseline	School roster at baseline, follow-up survey response (adjusted for time between surveys), imputation based on student age, and imputation based on the modal baseline grade within cluster
Race, Hispanic origin, and sexual attraction	Follow-up survey response

Note: Birth date, grade, and gender data were not included on rosters for all schools. In limited cases in which multiple imputation methods listed for the same variable above yielded contradictory answers, the study team gave preference to the source listed earlier.

Some student characteristics—race, parental education, and sexual attraction—had multiple contributing variables or a range of possible answers. To simplify the analysis, the study team also collapsed these variables into a shorter list of mutually exclusive categories, as shown in Exhibit A.7. Many outcome measures also were based on responses to multiple questions; this appendix addresses this issue later.

The study's benchmark analytic approach handled missing data in two ways, depending on whether the missing variable was an outcome or a baseline characteristic. When estimating impacts, the study team focused on individuals with observed outcome data for each outcome and excluded anyone with a missing response from analysis of the outcome in question. However, the observations included sometimes were missing key baseline data as well. Rather than excluding study participants with missing baseline data (that is, listwise deletion), the team used a dummy variable adjustment to address missing baseline data (Puma et al. 2009). Specifically, the team imputed missing data to a constant and included in its regression specifications an indicator variable for each baseline variable that had any missing data. This indicator variable was equal to one for participants whose baseline data were missing before imputation and zero for participants whose baseline data were not missing before imputation.

To ensure that the study's results were not driven by its approach to handling missing data, the study team took two other approaches to handling missing covariate and outcome responses. The first was a complete case analysis that excluded from analysis any observation missing one or more baseline characteristics, along with those missing outcome data. The second approach used a method called "multiple imputation." This method uses nonmissing response data to predict the values of missing

variables (Rubin 1987). For both outcome and control variables, the study team dealt with missing responses (including for individuals who responded to only one of the two surveys) using multiple imputation by chained equations. The team used an imputation algorithm called predictive mean matching, or PMM. For an observation missing the value of some variable Z, PMM uses nonmissing data for other variables to find several similar-looking observations with nonmissing values of Z, and then randomly chooses one of them to "donate" its value to the first observation. The process is repeated several times—10 times, in this case—with a new set of imputed values generated at random each time to account for uncertainty about the imputed values. The standard errors of the estimated impacts are then adjusted to account for the artificially increased sample size.

Exhibit A.7. Final set of mutually exclusive categories for student background data

Student characteristics	Mutually exclusive categories in cleaned responses
Race	• Black • White • Any other single race • Multiple races
Mother's and father's education	• High school diploma or less • Some college or more • Not known (but mother/father figure present) • Mother/father figure not present
Gender identity	• Male • Female • Student indicates some other gender identity
Sexual attraction and gender identity	• Male or female gender and attracted all/mostly to the opposite sex • Male or female gender and attracted all/mostly to the same sex • Male or female gender and attracted equally to both sexes • Male or female gender and some other response • Student indicates some other gender identity

Finally, to simplify interpretation of its results, the study team addressed a number of inconsistent student responses. Inconsistencies might arise, for example, if a student indicated at baseline that they had previously been sexually active but responded at follow-up that they had *never* been sexually active. For another example, at follow-up a student might indicate that in the past three months they had both (1) had vaginal sex five times, but (2) had vaginal sex *without a condom* 10 times (even though the second number should logically be no larger than the first).

To avoid favoring one answer over another when data were inconsistent, the study team treated responses for both questions that gave rise to the inconsistency as missing in its benchmark analysis. Moreover, the team also treated closely related responses as missing. For example, if a student gave inconsistent answers about the number of total and unprotected vaginal sex encounters in the past three months, all other follow-up survey answers pertaining to vaginal sex in the past three months would be set to missing, including responses for the number of partners and number of times without birth control other than condoms. As a result of this approach, 6 percent of nonmissing sexual behavior or health goal outcomes in the raw data were recoded or set to missing due to inconsistencies in contributing measures. The study team also conducted a robustness check that analyzed students' raw responses as given, even if logically related responses contradicted one another. The direction and significance of these findings were nearly identical to the benchmark analysis.

H. Measures of contrast between treatment and control experiences

The survey contained two variables that measured exposure to TPP information (Exhibit A.8). These variables were used to describe the contrast in services being received by students in the MPC program and those in the control group. They included (1) exposure to TPP and sexual health programming and (2) receipt of reproductive health information in a health care setting. These items were adapted based on items from the impact evaluation surveys used in the PREP Multi-Component Evaluation.

Exhibit A.8. Measures of exposure to TPP information

Measure	How the measure was defined
Exposure to TPP programming (α = 0.84)	Multiple-item continuous scale variable: proportion of TPP program topics covered in classes or sessions that the respondent attended in the past six months; variable ranges from 0 to 1, with higher values indicating exposure to more TPP programming topics. These topics include the following: • Relationships, dating, or marriage • Abstinence from sex • Methods of birth control, such as condoms, pills, the patch, the shot, the ring, IUD, or an implant • Where to get birth control • Sexually transmitted diseases, also known as STDs or STIs
Receipt of reproductive health information in a health care setting (α = 0.81)	Multiple-item continuous scale variable: proportion of reproductive health information that the respondent received in the past six months from a doctor, nurse, or clinic; variable ranges from 0 to 1, with higher values indicating receipt of more reproductive health topics. These topics include the following: • Methods of birth control, such as condoms, birth control pills, the patch, the shot, the ring, IUD, or an implant • Where to get birth control • Sexually transmitted diseases (STDs or STIs) • The HPV vaccine, also known as Gardasil or Cervarix

Note: If respondents left more than half of the items missing within each scale, the corresponding outcome measure was coded as missing.

When comparing the treatment and control groups on these variables, the study team found large, statistically significant differences in experiences observed across the two groups (Exhibit A.9).

Exhibit A.9. Observed impacts on differences in information observed across treatment and control groups at follow-up

Outcome	MPC mean	Control mean	Estimated impact	Effect size (ES)	*p*-value	MPC sample size	Control sample size
Exposure to TPP programming	0.60	0.41	0.19***	0.49	< 0.001	877	993
Receipt of reproductive health information in a health care setting	0.33	0.28	0.06*	0.16	0.014	805	887

Source: Six-month follow-up survey.

Note: Treatment and control group means are regression adjusted. All regressions adjust for strata fixed effects and after clustering standard errors at the school*cluster level. All *p*-values are based on a two-sided test. Sample sizes differ across outcomes due to missing data.

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

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

*** Significantly different from zero at the 0.001 level, two-tailed test.

I. Outcome measures

The study examined all outcomes described in the logic model for MPC (presented in Exhibit II.2), including short-term outcomes proximal to the program and more distal behavioral outcomes.

1. Risk and protective factors

The logic model in Exhibit II.2 highlights that knowledge, beliefs, attitudes, skills, and self-efficacy are risk and protective factors, which are precursors to behavioral change and the outcomes most proximal to the content of the MPC program. Individual bullets in the logic model represent specific constructs of interest associated with each risk and protective factor. The following tables describe how the study team operationalized each of these constructs using items from the follow-up survey.

The survey contained a series of true/false items, grouped into blocks around knowledge of HIV/STIs, pregnancy, condoms, and other birth control methods. Items in the HIV/STI knowledge block were adapted from survey items used in the original MPC evaluation (Jemmott et al. 1998); other knowledge items were adapted from those used on Power to Decide's Fog Zone Survey (Kaye et al. 2009) with a young adult survey population. Items used for the measures of beliefs, attitudes, and skills and self-efficacy were adapted from survey items used in the original MPC evaluation (Jemmott et al. 1998) and items from the evaluation surveys used in the PREP and the Evaluation of Adolescent Pregnancy Prevention Approaches, both conducted by Mathematica. Exhibit A.10 provides details about the construction of each measure and the specific item wording within each block, along with Cronbach α internal consistency statistics for each scale.

Exhibit A.10. TPP knowledge, beliefs, attitudes, skills, and self-efficacy measures

Outcome	Measure
Knowledge of:	
HIV/STIs (α = 0.78)	Proportion of knowledge items about HIV/STIs answered correctly: • The HIV virus is present in blood, semen, and vaginal fluid. (TRUE) • A person with HIV or AIDS can give it to other people only if they look or feel sick. (FALSE) • A woman who has an STD or STI can get an infection in her uterus and fallopian tubes. (TRUE) • You can have an STD or STI and feel healthy. (TRUE) • Birth control pills can reduce the risk of getting a sexually transmitted disease (STD). (FALSE) • Long-acting methods like the implant, an IUD, or the shot can reduce the risk of getting a sexually transmitted disease (STD). (FALSE) • Can you get a sexually transmitted disease, also known as an STD or STI, from having anal sex? (YES) • Can you get a sexually transmitted disease, also known as an STD or STI, from having oral sex? (YES)

Outcome	Measure
Pregnancy (α = 0.68)	Proportion of knowledge items about pregnancy answered correctly: • The only way to completely prevent pregnancy is by not having sex. (TRUE) • Pregnancy is much less likely to occur if a couple has sex standing up. (FALSE) • Douching (washing the vagina) after sex can prevent pregnancy. (FALSE) • The very first time a woman has sex, she cannot get pregnant. (FALSE) • During a woman's monthly cycle, there are certain days when she is more likely to become pregnant if she has sex. (TRUE)
Condoms (α = 0.70)	Proportion of knowledge items about condom use answered correctly: • Wearing two latex condoms will provide extra protection. (FALSE) • When using a condom, it is important for the man to pull out right after ejaculation. (TRUE) • It is okay to use petroleum jelly or Vaseline as a lubricant when using latex condoms. (FALSE) • When putting on a condom, it is important to leave a space at the tip. (TRUE) • Condoms have an expiration date. (TRUE) • It is okay to use the same condom more than once. (FALSE)
Other forms of contraception (α = 0.78)	Proportion of knowledge items about contraceptive use answered correctly: • It is not necessary for women to "take a break" from the pill every couple of years. (TRUE) • In order to get the birth control pill, a woman must have a pelvic exam. (FALSE) • After a woman stops taking birth control pills, she is unable to get pregnant for at least two months. (FALSE) • Birth control pills are less effective if a woman misses taking them for two or three days in a row. (TRUE) • An IUD is effective (prevents pregnancy) for at least three years. (TRUE) • An IUD cannot be felt by a woman's partner during sex. (TRUE) • A woman can get an IUD without going to a doctor's office, clinic, or medical professional. (FALSE) • Women who use IUDs cannot use tampons. (FALSE) • A woman can use an IUD, even if she has never had a child. (TRUE) • Long-acting methods like the implant or an IUD can make it more difficult to become pregnant in the future when a woman is no longer using them. (FALSE) • Long-acting methods like the implant (Implanon or Nexplanon) or IUD (Mirena, ParaGard, or Skyla) can be removed early if a woman changes her mind about wanting to get pregnant. (TRUE) • Women using the vaginal ring, NuvaRing, must have it inserted by a doctor or health care provider every month. (FALSE) • Women using the birth control shot, Depo Provera, must get an injection every three months. (TRUE)
Beliefs:	
That sex may adversely affect future goals (α = 0.71)	Proportion of statements about how sexual activity will interfere with goals and dreams that the respondent agrees or strongly agrees with: • Having sex at your age would make you less likely to have the career you are hoping for. • Having sex at your age would make you less likely to graduate from high school.
That condoms can be pleasurable (α = 0.70)	Proportion of statements about how condoms can be pleasurable that the respondent agrees or strongly agrees with: • Sex feels unnatural when a condom is used. [reverse code] • Condoms ruin the mood because you have to stop to put one on. [reverse code] • Condoms decrease sexual pleasure. [reverse code] • Condoms are embarrassing to use. [reverse code]

Outcome	Measure
Attitudes about:	
Condom use (α = 0.65)	Proportion of statements about condom use that the respondent either agrees or strongly agrees with: • Condoms are pretty easy to get. • Condoms are a hassle to use. [reverse code] • It is too much trouble to carry around condoms. [reverse code] • Condoms are important to make sex safer. • Condoms should always be used if a person your age has sexual intercourse.
Skill and self-efficacy related to:	
Condom use (α = 0.68)	Proportion of statements about self-efficacy to use condoms that the respondent agrees or strongly agrees with: • I am sure that I can use a condom if I have sex. • I feel confident I could put a condom on myself or my partner. • If I were "turned on," I could stop before sex to use a condom.
Condom use negotiation[a] (α = 0.75)	Proportion of statements about self-efficacy to negotiate condom use that the respondent agrees or strongly agrees with: • I could get my partner to use a condom, even if they didn't want to. • I could not put a condom on my partner or myself without ruining the mood. [reverse code] • I could get my partner to agree to use a condom without turning them off. • I feel confident talking about condom use. • I can say to my partner that we should use a condom. • I could say "no" to having sex if I was with someone who didn't want to use a condom.
Refusal (α = 0.45)	Proportion of statements about self-efficacy to refuse sex with a partner that the respondent agrees or strongly agrees with: • If I didn't want to have sex and was with someone who was pushing me to have sex, I could say "no." • If my dating partner wanted to have sex but I didn't, I would find it pretty hard to say "no." [reverse code]

Note: Answers of "don't know" were counted as incorrect responses for knowledge items. If respondents left more than half of the items missing within each scale, the corresponding outcome measure was coded as missing.

[a] The study used a narrower construct on this item than stated in the logic model, which included negotiating abstinence. Negotiating abstinence was captured within refusal skills.

2. Sexual behavior outcomes

The study team also estimated impacts on sexual behavior outcomes, which were expected to be mediated by changes in youth risk and protective factors. Consistent with the sexual behavior goals in the logic model, the team examined sexual initiation, frequency of sex, frequency of unprotected sex, and consistent condom use. Exhibit A.11 highlights how the study operationalized individual constructs within these domains using items from the follow-up survey.

As noted in the description of the approach for estimating impacts in the following section, the study team explored the extent to which the effects of MPC differed by site, demographic subgroups, and sexual experience defined at baseline. By conducting analyses for both a full study sample and the subset of youth who were sexually active at baseline, the study addressed the key behavioral outcomes specified in the logic model in Exhibit II.2.

Exhibit A.11. Sexual behavior outcomes

Outcome	Measure
Sexual initiation	
Ever had sex	Binary variable representing whether the respondent reported having ever had vaginal, oral, or anal sex • Have you ever had vaginal sex? • Have you ever had oral sex? • Have you ever had anal sex?
Decreased frequency of sex[a]	
Sexual activity in the last three months	Binary variable representing whether the respondent reported having vaginal, oral, or anal sex in the three months before completing the follow-up survey • Now please think about the past three months. In the past three months, have you had vaginal sex, even once? • Now please think about the past three months. In the past three months, have you had oral sex? • Now please think about the past three months. In the past three months, have you had anal sex?
Sexual episodes in the last three months	Count variable representing the total number of times respondent had vaginal, oral, and/or anal sex in the three months before completing the follow-up survey • In the past three months, how many times have you had vaginal sex? • In the past three months, how many times have you had oral sex? • In the past three months, how many times have you had anal sex?
Number of vaginal sexual partners in the last three months	Count variable representing the number of partners for vaginal sex in the three months before completing the follow-up survey • In the past three months, how many different people have you had vaginal sex with, even if only one time?
Risk of STIs	
Risk for STI sex (sex without a condom or dental dam) in last three months	Binary variable representing whether respondent reported having any risk for STI sex (sex without a condom or dental dam) in the three months before completing the follow-up survey • In the past three months, how many times have you had vaginal sex without you or your partner using a condom? • In the past three months, how many times have you had oral sex without using a condom or a dental dam? • In the past three months, how many times have you had anal sex without using a condom?
Risk for STI sex episodes in last three months	Count variable representing the total number of times respondent had risky sex in the three months before completing follow-up survey • In the past three months, how many times have you had vaginal sex without you or your partner using a condom? • In the past three months, how many times have you had oral sex without using a condom or a dental dam? • In the past three months, how many times have you had anal sex without using a condom?

Outcome	Measure
Risk of pregnancy	
Unprotected vaginal sex in the last three months	Binary variable representing whether the respondent reported having vaginal sex without birth control in the three months before completing the follow-up survey • The next question is about your use of the following methods of birth control: condoms, birth control pills, the shot, the patch, the ring, IUD, or implant. • In the past three months, how many times did you have vaginal sex without you or your partner using any of these methods of birth control?
Frequency of unprotected vaginal sex in last three months	Count variable representing the total number of times the respondent had unprotected vaginal sex in the three months before completing the follow-up survey • The next question is about your use of the following methods of birth control: condoms, birth control pills, the shot, the patch, the ring, IUD, or implant. • In the past three months, how many times did you have vaginal sex without you or your partner using any of these methods of birth control?

[a] The study team operationalized this outcome measure slightly differently from the logic model. The logic model specifies a decrease in frequency of sexual activity among sexually experienced youth. The team examined the impact on decreased frequency of activity for the full study sample to maximize the sample and power for this outcome. The team also estimated impacts on decreasing frequency of sex among youth who were sexually experienced at baseline as a key subgroup of interest.

3. MPC health goals

The logic model in Exhibit II.2 shows two health goals for MPC that are adversely affected when youth engage in risky sexual behaviors: (1) prevention of teen pregnancy and (2) prevention of HIV and other STIs. The study team operationalized these two outcome measures, as shown in Exhibit A.12.

Exhibit A.12. MPC health goals

Outcome	Measure
Pregnancy	
Pregnancy	Binary variable representing whether the respondent reported a pregnancy or has gotten someone pregnant • To the best of your knowledge, are you currently or have you ever been pregnant, or have you ever gotten someone pregnant?
STIs (including HIV)	
Sexually transmitted diseases/infections	Binary variable representing whether the respondent indicated having been told they have a sexually transmitted disease • Now think about the past three months. In the past three months, have you ever been told by a doctor, nurse, or some other health professional that you had any of the following sexually transmitted diseases (STDs or STIs)? – Chlamydia – Gonorrhea – Genital herpes – Syphilis – HIV infection or AIDS – Human papilloma virus, also known as HPV or genital warts – Other STD

J. Approach for estimating impacts

The study's analytic approach for estimating impacts was based on a sequence of four steps intended to produce face valid and credible findings: (1) assess baseline differences in the analytic samples, (2) address baseline differences in the impact estimation using regression controls, (3) calculate benchmark impact estimates, and (4) assess the sensitivity of the study's benchmark impact estimates to alternative assumptions. The study team also estimated impacts of MPC for key subgroups as a fifth and final step, following these same procedures. The following sections describe each of these steps in greater detail.

1. Assessing baseline equivalence

Random assignment designs create treatment and control groups that are similar, on average, on both observed and unobserved characteristics at the time of random assignment. Any observed difference among the samples at the time of random assignment can be attributed to random sampling error and is not a potential source of bias in an observed impact estimate. However, nonresponse at follow-up can potentially produce compositional differences in the two groups, contributing to the estimate of program effectiveness (that is, the analytic sample), which would lead to a biased test of the effect of the program.

The study team used data from the baseline survey to assess differences between the two groups that contributed to the estimates of program characteristics on observable characteristics, examining both the magnitude and the statistical significance of differences between the treatment and control groups. The team assessed baseline differences in participant characteristics expected to be associated with outcomes and baseline measures of each outcome of interest. Regarding participant characteristics, the team examined race, Hispanic origin, age, gender, parental presence, sexual orientation, and baseline risk behaviors, including school truancy, smoking, alcohol/drug use, sexual initiation, and relationship status. The team reported the baseline means and standard deviations, the difference in means in raw units and standard deviation (effect size) units, and a *p*-value for the inferential test of the difference.

Because this was a cluster randomized trial, all inferential analyses, including both analyses of baseline differences and program impacts, accounted for the assignment of groups of individuals. To accomplish this, the study team estimated a mixed-effects model that included school*cohort random effects, plus fixed effects for strata to account for assignment of schools to treatment or control conditions within a given city and cohort. The school*cohort random effects effectively captured the 31 clusters randomly assigned to condition.

As noted above, the follow-up survey was the source of data used to show MPC's effectiveness. There were slightly different analytic samples for each outcome due to item nonresponse in that follow-up survey; however, most individuals responded to most outcome measures. For the purposes of assessing baseline equivalence of the various analytic samples, the study team created an indicator variable representing whether an individual was a respondent for at least one outcome in the (1) risk and protective factor bucket, (2) sexual behavior bucket, or (3) MPC health goal bucket. The team assessed baseline equivalence for each of these analytic samples that differed slightly due to item nonresponse (see Appendix B for baseline equivalence findings).

2. Addressing baseline differences

To purge the small differences from the observed impact estimates that were due to the minor differences in characteristics observed at baseline, the study team adjusted for several categories of variables: all demographic characteristics, baseline risk variables, and baseline measures of each outcome variable in the study's benchmark analyses.[19] The team also adjusted for the length of time between baseline and

[19] The baseline survey did not measure belief that sex may adversely affect future goals or condom use self-efficacy. In the study's impact analyses, the study team adjusted for proxies for these outcomes: beliefs that condoms can be pleasurable and condom negotiation self-efficacy, respectively.

follow-up survey assessments, as well as indicators for whether there was a summer between assessment points, to address imbalances in these variables. In addition to mitigating small underlying differences in the groups that could act as potential sources of bias on the observed impact estimate, this approach also improved the precision of the impact estimate. This analytic approach of statistically adjusting for small differences in demographics and baseline measures (those less than 0.25 standard deviations) was a recommended one for benchmark/main impact analyses (Kautz and Cole 2017).

3. Estimating program impacts

The benchmark analysis this study used to estimate the impacts of MPC focused on individuals with observed (nonmissing) outcome data; as noted above, it statistically adjusted for several baseline and strata variables to produce credible and precise estimates of program effectiveness.

The study team estimated impacts using the following equation:

$$(1) \qquad y_{isc} = \alpha + \tau T_{sc} + X_{isc}\beta + W_{sc}\gamma + \mu_{sc} + \epsilon_{isc}$$

where y_{isc} was the outcome for individual i in school *s* and cohort *c*; T_{sc} was an indicator equal to one for participants whose school and cohort was assigned to the treatment group and zero for those assigned to the control group; X_{isc} was a vector of individual-level covariates (including the missing characteristic indicators described above); W_{sc} was a vector of stratum fixed effects used to account for how schools were randomly assigned to condition (for example, within a city*cohort, or serving as its own counterfactual across cohorts); μ_{sc} was a school-by-cohort-level random effect with an expected value of zero to capture systematic differences across randomized clusters; and ϵ_{isc} was an individual-level error term. The variables α and τ and the vectors β and γ were parameters to be estimated, where τ is the average treatment effect of having one's school and cohort assigned to MPC.

For each outcome examined, the study team reported the regression-adjusted means for the treatment and control groups, the impact estimate, the standardized effect size, the *p*-value from a two-sided test, and the posterior probability of the contrast. The team did not conduct any traditional multiple comparison adjustments as a means to address Type I errors; the interpretation of Bayesian probabilities (described below) provided the protection against spurious false positives.

4. Assessing sensitivity of impact estimates

To assess whether the impacts were an artifact of methodological choices, the study team conducted several sensitivity analyses. The team examined how its findings changed according to six alternative but credible methodological approaches to five issues: (1) handling inconsistent data, (2) imputation of missing data, (3) accounting for design fixed effects, (4) covariate adjustment, and (5) survey timing.

Handling inconsistent data. As noted above in the section that describes how data were prepared, this study's benchmark approach treated all inconsistent survey responses as missing (for example, individuals who indicated they were sexually active at baseline and then at follow-up indicated they were sexually inexperienced). As a sensitivity analysis, the study team analyzed the raw survey responses without treating inconsistent data as missing.

Imputation of missing data. This study's benchmark analysis used complete case analysis of outcome data after logical imputation of several variables (based on additional available data) and dummy variable imputation of missing baseline covariates. The study team conducted two sensitivity analyses around this issue. The first used complete case analysis of both outcome *and* baseline covariates, dropping any observation with any missing value among the variables included in the analytical model. In its second approach, the team multiplied imputed all missing covariate and outcome data among survey respondents—a process that algorithmically produced 10 versions of the full data set, with slight differences across them to account for uncertainty about missing data. This approach was capable of handling cluster-randomized data such as those in this project and produced a single analytic file with

complete data for all baseline and outcome measures. In doing so, the study's multiple imputation approach produced a single set of impact findings with a constant sample size, rather than one that varied across outcomes due to item nonresponse at follow-up. (Although multiple imputation increases the sample size for these analyses, the imputed values were obtained in such a way as to avoid artificially inflating the precision of the estimates. The study team also adjusted the standard errors to account for the multiple data sets produced.) The imputation approach included an indicator variable for condition, all but three of the baseline variables used as covariates in the impact model, and all outcomes—an approach expected to meet the current WWC Version 4.1 standards.[20]

Design fixed effects. The study's benchmark analysis accounted for the assignment of schools to treatment or control within a given cohorts-city pair by including a stratum fixed effect.[21] To test the sensitivity of findings to this approach, the study team conducted an analysis in which it excluded these fixed effects. Doing so sacrificed the precision gain by eliminating these dummy variables as predictors, but potentially offset this precision loss by gaining degrees of freedom in the analysis.

Covariate adjustment. As a sensitivity analysis, the study team estimated a parsimonious test of MPC by eliminating covariates, rather than adjusting for the baseline assessment of the outcome, several demographic and baseline risk characteristics, and the survey timing measures. The intent of this analysis was to estimate a crude, unadjusted effect of MPC to see the role the covariate adjustment played in the magnitude and significance of the observed effect.

Survey timing. The benchmark analytic approach statistically adjusted for the length of time between baseline and follow-up, as well as for an indicator variable for whether there was a summer between survey assessments (in addition to several other variables described earlier). As a sensitivity analysis, the study team eliminated the two survey timing covariates from model specification to understand the extent to which differences in survey timing contributed to the observed impact estimates.

The main report documents the number of times these sensitivity analyses produced findings that aligned with the direction and significance of the benchmark findings. This information was intended to help the reader understand the extent to which the findings were robust across different defensible analytic approaches (see Appendix C for sensitivity findings).

5. Subgroup analyses

In addition to testing the overall effect of MPC on the full study sample, the study team estimated impacts for several subgroups of interest. Such an analysis allowed for an exploration of heterogeneity in impacts across individuals and cities. This analytic approach largely mirrored those used for the full sample findings, but it calculated separate impacts for groups based on the following criteria:

Baseline sexual initiation status. One of the key findings from the original Jemmott et al. (1998) article was that the program was particularly effective for youth who were sexually active at baseline. Therefore,

[20] To successfully execute the multiple imputation approach, the study team was forced to exclude three baseline indicator variables from the analysis: (1) nonbinary gender, (2) sexual attraction responses such as "other" or "questioning," and (3) other race. These indicators represented rare characteristics in the sample; their inclusion in the imputation approach prevented the study's matching algorithm from finding values to impute for other missing variables, which is why the team dropped them in the final multiple imputation approach. Although the team excluded these characteristics from the set of imputed variables, due to the categorical nature of the study's gender, attraction, and racial characteristics, the analysis of the multiple imputation sample can still account for these characteristics in an equivalent way to the benchmark analysis.

[21] In fact, the study's stratification scheme was somewhat more complex than randomizing within city and cohort. For instance, the Cincinnati school was randomly assigned in the first cohort and switched to the other condition in the second cohort, placing it in its own single-school, cross-cohort stratum. The study's stratification variables reflect the actual approach used to randomize schools to condition in each cohort, and the sensitivity analyses around the use of stratification variables do not substantively affect the findings from the study.

the study team estimated impacts separately for youth who were sexually active at baseline and those not yet sexually active at that time.

Baseline gender. One goal of the study was to examine the extent to which the effect of MPC varied across youth with different demographic backgrounds. However, the sample ended up being largely homogeneous regarding youth age, race, and ethnicity, so there was little opportunity to explore several of the subgroups originally of interest. As a result, the study team examined heterogeneity of program impacts only across male and female gender subgroups (sample sizes were too small to examine any additional gender categories).

Location subgroups. The study team estimated and reported impacts separately for the study cities with at least four schools per city (Mobile and St. Louis). This set of results enabled the team to understand whether the program was more effective in one of these two settings.

Instance of random assignment. Nearly all 15 schools in the study participated in the random assignment lottery multiple times to enable a larger number of clusters—31 altogether—to be randomly assigned to condition and improve the statistical power of the evaluation. A potential limitation of that approach was that after the first instance of random assignment, youth previously assigned to the treatment group might affect the youth currently serving as the control group (for example, through dating), which had the potential to attenuate program impacts. Alternatively, health educators might feel more comfortable delivering the program in their second instance of delivery, which could magnify observed impacts. The study team estimated the effect of MPC for schools in their first instance of random assignment and for schools in subsequent rounds (for example, in their second or third rounds of participation).

Given that these subgroup analyses were calculated for a subset of the full study sample, the study had markedly less power to detect statistically significant impacts for any individual subgroups. That being said, understanding the extent to which the point estimates varied across subgroups (despite the imprecise standard errors) is useful in helping illustrate the heterogeneity of MPC's effectiveness.

6. Framework for understanding impact estimates

Researchers and decision makers know that some evaluation findings are more believable than others, but sorting out which findings deserve special attention can be challenging. Statistical significance is often used to determine which findings deserve attention, but the meaning of statistical significance is commonly misinterpreted. When an evaluation reports a statistically significant impact estimate, it is often misinterpreted to mean that there is a very high probability (for example, 95 percent) that the program had a positive effect.[22] This type of misinterpretation is so widespread that in 2016, the American Statistical Association issued a statement on the subject (Greenland et al. 2016; Wasserstein and Lazar 2016). The framework the study team recommends using is consistent with the association's guidance.

For this evaluation, the study team interpreted its impact estimates by calculating the probability that MPC truly had a positive effect, given the estimated impact. This type of probability is called a Bayesian posterior probability. To calculate this probability, the team used existing evidence on how often similar programs have positive effects in a given domain, commonly referred to as the "prior distribution" of potential effects. Based on previous evidence, the team expected to see relatively large effects on outcomes in the risk and protective factor domains but smaller impacts on sexual behaviors and the MPC health goals of preventing pregnancy and STIs. The Bayesian interpretation of MPC used evidence sources that reflected this nuance: the WWC for the domain of risk and protective factors (mean = 0.16,

[22] The *p*-value is the probability of estimating a program impact of the magnitude observed in the study (or larger) if the true impact is zero (that is, if the program does not actually change participant outcomes relative to the comparison group).

variance = 0.29; from Appendix D of Herrmann et al. 2019) and a meta-analysis of TPP programs for sexual behaviors and MPC health goal domains (mean = 0.03, variance = 0.15; from Juras et al. 2019).[23]

The study team recognizes that its approach to interpreting findings is not currently well known or widely used. To ensure that the findings were accessible to a broad audience, the team also used the traditional approach, in which impact estimates are presented with an indicator for whether they were significantly different from zero at the 0.05 level using a two-tailed test.

Sensitivity analysis around Bayesian interpretation. The standard approach to BASIE described in Deke and Finucane (2019) was geared towards interpreting one (or a small number) of impact estimates from a study. The MPC evaluation, however, has impacts on *20 outcomes*. Half of those outcomes are more proximal to the intervention, consisting of attitudes, beliefs, and knowledge. The other half are more distal (longer term) behavioral outcomes.

The large number of outcomes in the MPC evaluation created an opportunity to enhance the BASIE approach when interpreting each impact estimate. Instead of interpreting each MPC impact estimate using *only* evidence from prior studies, the study team also considered the impacts on all the other outcomes within the MPC evaluation itself. This amounted to a Bayesian meta-regression analysis of the MPC impact estimates. The role of the prior literature was to provide a broader context for interpreting the overall impact of MPC. Essentially, the team 'shrunk' each MPC impact estimate back to the MPC-wide mean, and then 'shrunk' the MPC-wide mean back to the external literature. The study team conducted two separate meta-regression analyses—one for the proximal outcomes and one for the distal outcomes.

Bayesian meta-regression is a 'textbook' methodology (Gelman et al. 2013), but applying the method to multiple impact estimates from the same study involved a complication that is often not encountered (or ignored) in typical meta-analyses. The complication was that all the impact estimates are *correlated*. To account for this correlation, the study team specified a *joint* distribution for the MPC impact estimates in the Bayesian meta-regression analysis (specifying the impact estimates as following the multivariate t-distribution). The team used bootstrapping to estimate the correlation matrix for this joint distribution. Because this correlation matrix was itself an estimate that was subject to uncertainty, the team used the weakly informative prior distribution described by Lewandowski, Kurowicka, and Joe (2009), which Gelman et al. (2013) recommends to regularize estimated correlation matrixes. The study team validated this method for its context using simulations.

This fully Bayesian approach to calculating posterior probability statistics is presented as a sensitivity analysis in Appendix C.

K. Implementation study data sources

The study of MPC implementation examined program delivery from April 2017 through May 2019. The implementation study relied on the following qualitative and quantitative data sources:

- Fidelity and attendance logs completed by health educators
- Observations of implementation and health educator training
- Summaries of TA provided by HTN
- Interviews with school staff, health educators, and site supervisors
- Student focus groups
- Survey of MPC health educators

[23] Means and standard deviations from both prior distributions are in standard deviation units for the outcome in question, as observed in the control group.

- Student surveys

1. Fidelity and attendance logs completed by health educators

All MPC health educators completed a fidelity and attendance log through an online form for each lesson they implemented. The log included multiple-choice questions and open-ended fields for health educators to provide more details on their implementation experiences. In total, the health educators completed 1,050 fidelity and attendance logs.[24]

The logs assessed several features of implementation. Health educators first noted which students with consent to participate in the study were present during that class period. Next the educators answered questions on fidelity. They noted whether they fully completed, partially completed, or did not complete at all each of the activities in the lessons. When they were unable to complete an activity in part or at all, they documented what they were unable to complete and why. In addition, health educators documented whether they made any additions or changes to the curriculum. Finally, they indicated whether they felt very prepared, somewhat prepared, or not prepared at all for delivery of that lesson.

2. Observations of implementation and health educator training

During the 2018–2019 and 2019–2020 school years (Cohorts 2 and 3), the study team conducted 40 observations covering 53 MPC lessons (approximately 5 percent of all implemented lessons).[25] Before the observations, the team trained data collectors on the observation form, including metrics and benchmarks for assigning quality measures. The study team purposefully oversampled the key lessons identified through discussions with the developer and HTN (Lessons 8 and 11 through 14) for observation so the skill-building activities were observed more frequently. In each of the four study sites, the team observed at least four lessons. Overall, the data from the observations may not be representative of all lessons implemented, but they provide some insights into the experiences implementing the lessons considered most important for improving behavioral outcomes.

Each observation of program delivery used the same 12-question rubric for assessing lesson quality. The rubric used a 1 to 5 scale, with 1 indicating poor quality, 3 average quality, and 5 excellent quality. The rubric provided a description and operationalization of what poor, average, and excellent quality would look like, tailored to the specific topic of that question. For instance, on the metric of classroom quality, the rubric clarified that poor quality would include the health educator not having control of the classroom, with students not focused on activities and classroom management issues consuming the teaching time. Average quality on this metric would include the health educator displaying some classroom management skills, keeping students on task for most of the lesson, but needing to address some classroom management issues taking up time. Finally, the health educator would display excellent quality on this metric if they managed the class well and students remained on task and focused throughout the lesson.

The study team observed two of the four MPC trainings.[26] In April 2017, a study team member observed the two-day MPC training an HTN staff member provided to five health educators in Mobile. In September

[24] In one quarter in Cohort 2 in Mobile, health educators did not complete logs. The study team collected attendance data for consented students from the schools but did not have data on fidelity for this quarter.

[25] Some schools had block scheduling that allowed for two lessons to be completed in a single class period. In addition to these 40 observations, the team conducted five observations (covering nine lessons) during Cohort 1 in spring 2017. However, the study team did not include data from these Cohort 1 observations in its analysis because the team identified inconsistencies in the scoring of observations; it was determined that the observers needed additional training to score the quality of implementation consistently. Between Cohorts 1 and 2, the study team retrained the observers, allowing for more consistent observation data in Cohorts 2 and 3.

[26] The study team did not observe training in Cincinnati or in Cohort 2 in Detroit, but it did hold debrief conversations with the HTN trainer to hear her perspective on how training went. Separately, HTN held a booster training for the Mobile health educators before the start of Cohort 2. This half-day booster took place in person.

2018, another study team member observed the two-day health educator training in St. Louis. Two HTN staff members conducted this training for 13 participants—11 from St. Louis and two from Detroit.

3. Summaries of technical assistance provided by HTN

The TA providers from HTN summarized the TA provided to health educators in logs completed one to three times per cohort for each site. In all, HTN completed 10 logs: one for Cincinnati, two for Detroit, four for Mobile, and three for St. Louis. The logs included information on the type of contact (email, individual phone call, group phone call), duration, health educators involved, and a summary of the discussions.

4. Interviews with school staff, health educators, and site supervisors

During six site visits, the study team collected data for the implementation study from health educators, site supervisors, and school staff from all 15 study schools. At each school, the study team interviewed two to four staff members. Interviews typically included the classroom teacher(s) whose students were in the evaluation, the principal, or a school guidance counselor or other school staff person who had been involved in the study. The team conducted all but one interview with a single respondent; in one case, the study team interviewed two school staff members together. In all, the team completed 65 interviews with 56 total respondents, as the study team interviewed some people multiple times (Exhibit A.13).[27]

Across the interviews, the team collected in-depth data on the following:

- Staff preparation and receipt of training and TA
- MPC implementation (when assigned to MPC)
- School and student characteristics
- Other similar programming offered to students in the school or community
- Experiences of control group youth, including content covered in their health classes
- Lessons learned from program implementation

Exhibit A.13. Number of interviews, by study condition and respondent type

Respondent type	MPC	Control	Total
Principal/assistant principal/superintendent	7	7	14[a]
Classroom teacher	12	9	21[b]
Other school or district staff	4	4	8[c]
Health educator	15	3	18[d]
Site supervisor	4	n.a.	4[e]
Total	42	23	65

Note: Study team members conducted interviews in person during site visits, except for six interviews conducted over the phone. The team interviewed several individuals, multiple times. One interview with other school staff included two people. In all, 56 respondents completed at least one interview with the study team.

[a] One principal was interviewed twice—once when the school was assigned to MPC and once when it was assigned to control.

[b] Three classroom teachers were interviewed twice. In two cases, the study team interviewed the teachers once when the school was assigned to MPC and once when it was assigned to control.

[c] One staff person was interviewed twice—once when the school was assigned to MPC and once when the school was assigned to control.

[27] In six cases, the study team conducted interviews over the phone.

[d] Four health educators were interviewed twice because they implemented MPC at multiple schools or in multiple cohorts. Two of the MPC health educators interviewed in St. Louis also implemented Too Good for Drugs and Violence to the control group, so the study team asked them about their experiences implementing both curricula. These two health educators are counted only in the MPC column in the table. In all, the team interviewed 14 health educators.

[e] One site supervisor was interviewed twice across different cohorts. The site supervisor in St. Louis oversaw implementation of both MPC and Too Good for Drugs and Violence, so the study team asked this supervisor about the team's experiences implementing both curricula.

n.a. = not applicable.

5. Student focus groups

During site visits, staff conducted 12 focus groups with students in the treatment group. These groups included one focus group in Cincinnati during Cohort 2, two in Detroit during Cohorts 2 and 3, three in Mobile in Cohort 2, and six in St. Louis in Cohort 3. The focus groups ranged from four to 17 participants, with an average size of eight students. In all, 99 students participated in a focus group. The focus groups took place during or immediately following MPC implementation at the students' school, depending on the timing of the site visit and MPC schedule.

Focus group topics covered the following:

- Students' perceptions of MPC, including its materials, topics, and activities
- Suggested changes to MPC
- Participation and engagement in MPC, including any barriers to participation
- Receipt of other similar programming at school or in the community

During focus groups, students also discussed whether participating in MPC led them to feel differently about sexual health topics or change their behavior.

6. Survey of MPC health educators

Nine of the 11 MPC health educators completed a 30-minute survey of their experiences during MPC implementation. Health educators completed the survey once, after finishing implementation in their first cohort. The survey was in a pencil-and-paper format; health educators mailed their completed surveys to the study team. The survey collected data on staff characteristics and background, receipt of training and TA, perceptions of the training received, experiences with program implementation, and perceptions of MPC's fit for students' needs.

7. Student surveys

The study team relied on data from the student follow-up surveys for additional information on the TPP and sexual health programming students received. The discussion above on the impact analyses includes more information on this data source.

L. Implementation study analytic methods

The study team used a mixture of qualitative and quantitative methods to analyze implementation data on five topics: training and TA received by health educators, student attendance, implementation fidelity, implementation quality, and youth exposure to TPP and sexual health education and information. Descriptions of the analysis methods for each topic are below.

For qualitative data from interviews and student focus groups, the study team used an iterative process to identify key themes across different respondents (Patton 2002; Ritchie and Spencer 2002). In particular,

the team developed a coding scheme organized around the topics covered in the interview and focus group protocols. Next, the study team trained a research analyst and research associate on the coding scheme; they used the qualitative analysis software package NVivo to apply codes to passages from transcripts of the interviews and focus groups. To ensure accurate and consistent coding, a member of the site visit team conducted quality assurance reviews and provided feedback to the coders as needed. The team retrieved relevant passages from NVivo to examine the patterns of responses across respondents and identify common themes.

Training and TA received by health educators. The study team synthesized qualitative data from the study team's training observations and the HTN provider's TA summaries, and tabulated averages of responses from the staff survey.

Student attendance. Health educators collected attendance data in their logs (except for one quarter in Mobile in Cohort 2, when the participating schools provided attendance data). In some cases, students left class earlier or arrived late, so they were not present for the full class period. For the study, students were considered present if they attended any portion of the lesson.[28]

The sample for the attendance data analysis was limited to individuals in the impact study analytic sample for whom attendance data were available. Ten students in this sample did not have attendance data and so were not included in the analysis. The study team calculated the average attendance rate and the percentage of students who attended benchmarks (at least one, seven, 10, or 14 lessons and the five key lessons) for the impact analytic sample overall and by site.

Implementation fidelity. Using data entered by health educators in the fidelity and attendance logs, the study team calculated the percentage of lessons with all activities fully completed, with an addition to the curriculum, and with an adaptation to the curriculum, overall and by site. These calculations did not include the one quarter in Cohort 2 in Mobile during which health educators did not complete fidelity and attendance logs.

In their logs, health educators could note if they made up material not completed in an earlier lesson. Health educators completed some of the activities later than expected. Specifically, 7 percent of lessons had at least one activity not fully completed on the expected day; in most cases (58 percent of these lessons), the health educator completed the activities during a later class. Overall, health educators fully completed all activities on the expected day in 93 percent of lessons.

Using open-ended data in the logs and occasional follow-up email correspondence with health educators to gather more information, the study team worked with HTN to identify whether changes were green-, yellow-, or red-light adaptations. The team aligned these determinations with the MPC fifth edition adaptation guidelines published by ETR (2017). The team then calculated the percentage of adaptations that fell into each adaptation category.

Implementation quality. The study team calculated the mean scores on several questions from observations of MPC implementation; these questions were scale measures with scores ranging from 1 (poor) to 5 (excellent). The team also used relevant passages from youth focus groups, identified through the NVivo coding process, to assess quality.

Youth exposure to TPP and sexual health education and information. The study team tabulated information from the student survey and drew on qualitative data from interviews of control group staff.

[28] For the one quarter in Mobile for which attendance data came from schools, students were considered present for MPC if they were present for school that day; it is possible that students may have been at school but did not attend MPC class.

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

Appendix B: Baseline equivalence

The random assignment design of this study should, in expectation, produce treatment and control groups that were well balanced on all measured and unmeasured variables at the time of random assignment. To assess this situation, the study team compared the treatment and control groups on several characteristics and baseline measures of all outcomes of interest using the full study sample. Characteristics included age, grade, race, Hispanic origin, gender, sexual preference, parent influences, and behavioral risk factors. As Appendix A notes, the team also compared the treatment and control groups on the length of time between the baseline and the follow-up survey.

Although a few differences were statistically significant, a few differences caused by chance alone were to be expected when examining a large number of variables. Specifically, the control group, had slightly more youth of two or more races, and the treatment group had a longer time between baseline and follow-up than the control group. There were also some large baseline differences observed (greater than 0.25 standard deviations). There were more youth in 10th grade in the MPC group than the control group (and fewer 9th graders). In addition, the MPC group was more likely to experience a summer between baseline and follow-up. These differences largely stemmed from one school in Mobile that did not change conditions during the study (see Appendix A for more details). Aside from these differences, nearly all of the other variables examined showed differences that were small in magnitude; furthermore, the benchmark analytic approach adjusted for all variables shown in Exhibit B.1.

Exhibit B.1. Baseline characteristics for the full study sample

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Age and grade							
Age at baseline (years)	15.75	15.60	0.15	0.17	0.27	1,007	1,127
9th grade	0.34	0.48	-0.14	-0.28	0.23	1,010	1,128
10th grade	0.60	0.47	0.13	0.26	0.26	1,010	1,128
11th grade	0.04	0.04	0.00	-0.02	0.87	1,010	1,128
12th grade	0.02	0.01	0.01	0.07	0.30	1,010	1,128
Race and Hispanic origin							
Black	0.86	0.77	0.08	0.21	0.30	976	1,101
White	0.05	0.11	-0.06	-0.20	0.42	976	1,101
Other race	0.01	0.02	-0.01	-0.07	0.16	976	1,101
Two or more races	0.07	0.10	-0.03	-0.09	0.04	976	1,101
Hispanic	0.04	0.04	0.00	-0.02	0.78	987	1,114
Gender and sexual preferences							
Male	0.45	0.46	-0.01	-0.02	0.74	1,004	1,127
Female	0.53	0.52	0.01	0.02	0.67	1,004	1,127
Nonbinary	0.01	0.01	0.00	-0.02	0.76	1,004	1,127
Binary gender and attracted to opposite gender	0.83	0.81	0.02	0.05	0.36	974	1,106
Binary gender and attracted to same gender	0.05	0.05	0.00	0.00	0.95	974	1,106

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Binary gender and equally attracted to male and female genders	0.07	0.09	-0.02	-0.08	0.07	974	1,106
Binary gender and questioning/not sure who they are attracted to	0.04	0.03	0.00	0.03	0.57	974	1,106
Parent influences							
Presence of a father figure	0.77	0.76	0.01	0.01	0.79	942	1,075
Presence of a mother figure	0.96	0.96	0.00	-0.02	0.65	946	1,079
Mother's highest education was high school	0.29	0.30	-0.01	-0.02	0.65	1,010	1,128
Mother had some college or more education	0.36	0.38	-0.02	-0.05	0.51	1,010	1,128
Father's highest education was high school	0.29	0.28	0.00	0.01	0.89	1,010	1,128
Father had some college or more education	0.18	0.19	-0.01	-0.01	0.80	1,010	1,128
Behavioral risks (nonsexual) at baseline							
Number of times suspended or expelled	0.95	0.90	0.05	0.06	0.54	947	1,079
Ever skipped school	0.20	0.22	-0.01	-0.03	0.46	989	1,117
Proportion of substances used (vape, cigarettes, alcohol, marijuana, other illicit drugs)	0.11	0.12	-0.01	-0.02	0.73	871	1,008
Ever contacted someone online to meet up in person	0.28	0.28	0.00	0.00	0.98	872	999
Survey timing							
Days between baseline and follow-up surveys	277.66	243.36	34.30	0.42	0.01	831	957
Presence of a summer between baseline and follow-up	0.62	0.49	0.14	0.28	0.18	1,010	1,128
Risk and protective factor baseline assessments[a]							
Knowledge about HIV/STIs	0.41	0.41	0.00	-0.01	0.84	919	1,030
Knowledge about pregnancy	0.50	0.53	-0.02	-0.08	0.38	929	1,045
Knowledge about condoms	0.38	0.40	-0.02	-0.09	0.19	923	1,038
Knowledge about other forms of contraception	0.13	0.13	-0.01	-0.05	0.52	917	1,032
Belief that condoms can be pleasurable	0.14	0.14	-0.01	-0.02	0.78	910	1,028
Attitudes about condoms	0.68	0.66	0.02	0.06	0.17	927	1,041
Condom negotiation self-efficacy	0.82	0.83	-0.01	-0.02	0.77	930	1,046
Refusal self-efficacy	0.73	0.74	-0.01	-0.03	0.54	922	1,051

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Sexual behavior baseline assessments[a]							
Ever any sex (vaginal, oral, or anal)	0.32	0.36	-0.03	-0.07	0.23	897	1,022
Any sex in the past three months	0.22	0.21	0.01	0.02	0.75	897	1,018
Times having any sex in the past three months	1.55	1.81	-0.26	-0.03	0.62	892	1,011
Count of vaginal sex partners in the past three months	1.20	1.21	-0.01	0.00	0.95	864	974
Any sex without a condom in the past three months	0.15	0.14	0.01	0.03	0.57	893	1,003
Times having any sex without a condom in the past three months	0.73	1.04	-0.30	-0.04	0.39	893	1,003
Sex without birth control in the past three months	0.11	0.10	0.00	0.02	0.80	848	961
Times having sex without birth control in the past three months	0.29	0.70	-0.41	-0.10	0.07	848	961
MPC health goal baseline assessments[a]							
Ever pregnant	0.01	0.02	0.00	-0.01	0.82	855	961
Ever had an STI	0.09	0.09	0.00	-0.02	0.76	887	1,021

Sources: Baseline student survey, school rosters, and parent consent forms.

Note: Treatment and control means were model based, after adjusting for strata fixed effects. The *p*-values were based on an inferential test that adjusted the standard errors of the impact estimate for clustering. The baseline survey did not measure belief that sex may adversely affect future goals or condom use self-efficacy. In its impact analyses for these outcomes, the study team adjusted for proxies for these outcomes: belief that condoms can be pleasurable and condom negotiation self-efficacy, respectively. Missing data on the days between baseline and follow-up survey variable was predominantly due to individuals who completed a follow-up survey without completing the baseline.

[a] See Appendix A for details on how each outcome was operationalized.

Because item nonresponse on the follow-up survey created slightly different analytic samples for each outcome, the study team also examined the baseline equivalence of three versions of its eventual analytic samples, defined by the broad domains. This approach ensured that, as with the full sample, respondents who contributed to the impact estimates in each domain were reasonably well matched at baseline, and that no set of outcomes suffered from meaningfully disparate changes in the treatment or control groups. Exhibits B.2, B.3, and B.4 compare characteristics at baseline for treatment and control group members who provided survey responses to questions about risk and protective factors, sexual behaviors, and MPC health goals, respectively. The study team formed these three groups by grouping all students who answered *any* risk and protective factor questions (876 treatment students, 991 controls), *any* sexual behavior questions (835 treatment students, 934 controls), and *any* health goal questions (846 treatment students, 954 controls). These sample sizes reflect not only item nonresponse—decreases in sample size due to students' skipping certain questions—but also the data cleaning steps taken to drop inconsistent answers, as described in Appendix A.

The study team again found few statistically significant differences in any of these tables and, given the large number of variables examined, a few chance differences would be expected. Again, nearly all of the

differences shown here were small in magnitude; also, as noted above, the team adjusted for them in the benchmark analytic approach used to estimate program impacts. These three domain-specific samples exhibited baseline differences between the treatment and control groups similar to those of the main sample, with large or statistically significant differences occurring for the same set of characteristics.

Exhibit B.2. Baseline characteristics for respondents to risk and protective factor questions

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Age and grade							
Age at baseline (years)	15.73	15.57	0.16	0.19	0.20	876	991
9th grade	0.33	0.47	-0.14	-0.29	0.22	876	991
10th grade	0.62	0.48	0.14	0.28	0.22	876	991
11th grade	0.03	0.03	0.00	-0.02	0.82	876	991
12th grade	0.01	0.01	0.00	0.02	0.78	876	991
Race and Hispanic origin							
Black	0.85	0.77	0.08	0.20	0.33	864	976
White	0.05	0.11	-0.06	-0.19	0.43	864	976
Other race	0.01	0.02	-0.01	-0.06	0.31	864	976
Two or more races	0.08	0.10	-0.02	-0.09	0.06	864	976
Hispanic	0.04	0.04	0.00	-0.02	0.77	875	988
Gender and sexual attraction preferences							
Male	0.46	0.47	-0.01	-0.01	0.86	876	990
Female	0.53	0.52	0.01	0.02	0.75	876	990
Nonbinary	0.01	0.02	0.00	-0.03	0.72	876	990
Binary gender and attracted to opposite gender	0.83	0.82	0.02	0.04	0.44	869	985
Binary gender and attracted to same gender	0.05	0.05	0.00	-0.01	0.77	869	985
Binary gender and equally attracted to male and female genders	0.07	0.09	-0.02	-0.06	0.23	869	985
Binary gender and questioning/not sure who they are attracted to	0.04	0.03	0.01	0.03	0.52	869	985
Parent influences							
Presence of a father figure	0.78	0.77	0.01	0.02	0.75	831	949
Presence of a mother figure	0.96	0.96	0.00	-0.02	0.74	835	953
Mother's highest education was high school	0.30	0.30	0.01	0.01	0.80	876	991
Mother had some college or more education	0.36	0.39	-0.03	-0.07	0.39	876	991
Father's highest education was high school	0.30	0.29	0.02	0.03	0.56	876	991
Father had some college or more education	0.19	0.19	0.00	-0.01	0.82	876	991

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Behavioral risks (nonsexual) at baseline							
Number of times suspended or expelled	0.94	0.87	0.07	0.08	0.44	835	953
Ever skipped school	0.19	0.21	-0.02	-0.04	0.43	876	990
Proportion of substances used (vape, cigarettes, alcohol, marijuana, other illicit drugs)	0.11	0.11	-0.01	-0.03	0.67	772	895
Contacted someone online to meet up in person	0.27	0.28	-0.01	-0.02	0.69	775	885
Survey timing							
Days between baseline and follow-up surveys	277.58	243.86	33.72	0.41	0.01	829	948
Presence of a summer between baseline and follow-up	0.62	0.48	0.13	0.27	0.19	876	991
Risk and protective factor baseline assessments[a]							
Knowledge about HIV/STIs	0.40	0.42	-0.01	-0.04	0.57	813	907
Knowledge about pregnancy	0.50	0.53	-0.03	-0.10	0.30	822	921
Knowledge about condoms	0.37	0.40	-0.03	-0.11	0.15	818	914
Knowledge about other forms of contraception	0.12	0.13	-0.01	-0.06	0.44	813	909
Belief that condoms can be pleasurable	0.13	0.15	-0.02	-0.05	0.42	803	907
Attitudes about condoms	0.68	0.66	0.02	0.09	0.08	821	919
Condom negotiation self-efficacy	0.83	0.82	0.00	0.00	0.98	822	922
Refusal self-efficacy	0.73	0.74	-0.02	-0.05	0.35	815	926
Sexual behavior baseline assessments[a]							
Ever any sex (vaginal, oral, or anal)	0.30	0.34	-0.03	-0.07	0.25	789	898
Any sex in the past three months	0.20	0.20	0.00	-0.01	0.89	789	894
Times having any sex in the past three months	1.33	1.78	-0.45	-0.05	0.41	784	889
Count of vaginal sex partners in the past three months	0.86	1.22	-0.36	-0.08	0.15	759	852
Any sex without a condom in the past three months	0.14	0.14	0.00	0.00	0.98	785	882
Times having any sex without a condom in the past three months	0.71	1.10	-0.39	-0.05	0.33	785	882
Sex without birth control in the past three months	0.09	0.11	-0.01	-0.04	0.46	745	839
Times having sex without birth control in the past three months	0.25	0.77	-0.51	-0.12	0.05	745	839
MPC health goal baseline assessments[a]							
Ever pregnant	0.01	0.01	0.00	-0.01	0.77	751	839
Ever had an STI	0.09	0.09	0.00	0.00	1.00	785	903

Sources: Baseline student survey, school rosters, and parent consent forms.

Note: Treatment and control means were model based, after adjusting for strata fixed effects. The *p*-values were based on an inferential test that adjusted the standard errors of the impact estimate for clustering. The baseline survey did not measure belief that sex may adversely affect future goals or condom use self-efficacy. In its impact analyses for these outcomes, the study team adjusted for proxies for these outcomes: belief that condoms can be pleasurable and condom negotiation self-efficacy, respectively. Missing data on the days between baseline and follow-up survey variable was predominantly due to individuals who completed a follow-up survey without completing the baseline.

[a] See Appendix A for details on how each outcome was operationalized.

Exhibit B.3. Baseline characteristics for respondents to sexual behavior questions

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Age and grade							
Age at baseline (years)	15.73	15.57	0.16	0.19	0.22	835	934
9th grade	0.33	0.48	-0.14	-0.29	0.22	835	934
10th grade	0.62	0.48	0.14	0.29	0.21	835	934
11th grade	0.03	0.03	-0.01	-0.04	0.68	835	934
12th grade	0.02	0.01	0.00	0.02	0.77	835	934
Race and Hispanic origin							
Black	0.85	0.77	0.08	0.19	0.34	824	919
White	0.05	0.11	-0.06	-0.19	0.43	824	919
Other race	0.02	0.02	-0.01	-0.06	0.29	824	919
Two or more races	0.08	0.10	-0.02	-0.08	0.10	824	919
Hispanic	0.04	0.04	-0.01	-0.03	0.71	834	931
Gender and sexual attraction preferences							
Male	0.45	0.46	-0.01	-0.03	0.67	835	934
Female	0.54	0.52	0.02	0.03	0.57	835	934
Nonbinary	0.01	0.02	0.00	-0.03	0.73	835	934
Binary gender and attracted to opposite gender	0.83	0.82	0.01	0.03	0.57	833	932
Binary gender and attracted to same gender	0.05	0.05	0.00	-0.01	0.89	833	932
Binary gender and equally attracted to male and female genders	0.07	0.09	-0.02	-0.06	0.21	833	932
Binary gender and questioning/not sure who they are attracted to	0.04	0.03	0.01	0.05	0.31	833	932
Parent influences							
Presence of a father figure	0.78	0.77	0.01	0.01	0.84	792	895
Presence of a mother figure	0.96	0.96	0.00	0.00	0.98	796	899
Mother's highest education was high school	0.30	0.29	0.01	0.02	0.70	835	934
Mother had some college or more education	0.36	0.40	-0.04	-0.07	0.33	835	934
Father's highest education was high school	0.30	0.28	0.01	0.03	0.62	835	934
Father had some college or more education	0.19	0.19	-0.01	-0.02	0.75	835	934
Behavioral risks (nonsexual) at baseline							
Number of times suspended or expelled	0.94	0.86	0.07	0.08	0.44	796	899
Ever skipped school	0.20	0.20	-0.01	-0.01	0.77	835	933

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Proportion of substances used (vape, cigarettes, alcohol, marijuana, other illicit drugs)	0.11	0.11	0.00	-0.02	0.79	741	846
Contacted someone online to meet up in person	0.26	0.27	-0.01	-0.02	0.75	743	839
Survey timing							
Days between baseline and follow-up surveys	278.12	244.32	33.80	0.41	0.01	794	896
Presence of a summer between baseline and follow-up	0.62	0.48	0.14	0.28	0.18	835	934
Risk and protective factor baseline assessments[a]							
Knowledge about HIV/STIs	0.41	0.42	-0.01	-0.03	0.70	779	858
Knowledge about pregnancy	0.50	0.53	-0.03	-0.09	0.31	785	870
Knowledge about condoms	0.37	0.40	-0.03	-0.10	0.18	783	864
Knowledge about other forms of contraception	0.12	0.13	-0.01	-0.04	0.58	779	860
Belief that condoms can be pleasurable	0.13	0.15	-0.01	-0.04	0.48	766	857
Attitudes about condoms	0.68	0.66	0.02	0.09	0.10	784	869
Condom negotiation self-efficacy	0.83	0.83	0.00	0.00	0.97	783	875
Refusal self-efficacy	0.73	0.75	-0.01	-0.04	0.46	776	876
Sexual behavior baseline assessments[a]							
Ever any sex (vaginal, oral, or anal)	0.30	0.33	-0.03	-0.07	0.25	775	869
Any sex in the past three months	0.20	0.20	-0.01	-0.01	0.82	775	865
Times having any sex in the past three months	1.33	1.82	-0.49	-0.05	0.37	770	861
Count of vaginal sex partners in the past three months	0.86	1.24	-0.38	-0.09	0.13	746	828
Any sex without a condom in the past three months	0.13	0.14	-0.01	-0.01	0.81	771	854
Times having any sex without a condom in the past three months	0.72	1.13	-0.41	-0.05	0.31	771	854
Sex without birth control in the past three months	0.09	0.11	-0.02	-0.06	0.31	732	814
Times having sex without birth control in the past three months	0.25	0.68	-0.43	-0.11	0.04	732	814
MPC health goal baseline assessments[a]							
Ever pregnant	0.01	0.01	0.00	-0.02	0.73	738	815
Ever had an STI	0.09	0.09	0.00	-0.01	0.91	752	852

Sources: Baseline student survey, school rosters, and parent consent forms.

Note: Treatment and control means were model based, after adjusting for strata fixed effects. The *p*-values were based on an inferential test that adjusted the standard errors of the impact estimate for clustering. The baseline survey did not measure belief that sex may adversely affect future goals or condom use self-efficacy. In its impact analyses for these outcomes, the study team adjusted for proxies for these outcomes: belief that condoms can be pleasurable and condom negotiation self-efficacy, respectively. Missing data on

the days between baseline and follow-up survey variable was predominantly due to individuals who completed a follow-up survey without completing the baseline.

[a] See Appendix A for details on how each outcome was operationalized.

Exhibit B.4. Baseline characteristics for respondents to MPC health goal questions

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Age and grade							
Age at baseline (years)	15.73	15.57	0.16	0.19	0.20	846	954
9th grade	0.33	0.48	-0.14	-0.29	0.21	846	954
10th grade	0.62	0.48	0.14	0.30	0.20	846	954
11th grade	0.03	0.04	-0.01	-0.03	0.70	846	954
12th grade	0.02	0.01	0.00	0.02	0.74	846	954
Race and Hispanic origin							
Black	0.85	0.77	0.08	0.19	0.33	835	939
White	0.05	0.11	-0.06	-0.19	0.44	835	939
Other race	0.02	0.02	-0.01	-0.06	0.30	835	939
Two or more races	0.07	0.10	-0.02	-0.08	0.08	835	939
Hispanic	0.04	0.04	0.00	-0.02	0.74	845	951
Gender and sexual attraction preferences							
Male	0.45	0.46	-0.01	-0.02	0.71	846	954
Female	0.54	0.52	0.02	0.03	0.57	846	954
Nonbinary	0.01	0.02	0.00	-0.04	0.62	846	954
Binary gender and attracted to opposite gender	0.83	0.82	0.01	0.03	0.52	845	951
Binary gender and attracted to same gender	0.04	0.05	-0.01	-0.03	0.58	845	951
Binary gender and equally attracted to male and female genders	0.07	0.09	-0.01	-0.04	0.40	845	951
Binary gender and questioning/not sure who they are attracted to	0.04	0.03	0.01	0.04	0.43	845	951
Parent influences							
Presence of a father figure	0.78	0.77	0.01	0.02	0.75	803	916
Presence of a mother figure	0.96	0.96	-0.01	-0.03	0.60	807	920
Mother's highest education was high school	0.31	0.30	0.00	0.01	0.84	846	954
Mother had some college or more education	0.36	0.39	-0.03	-0.07	0.36	846	954
Father's highest education was high school	0.30	0.29	0.01	0.03	0.64	846	954
Father had some college or more education	0.18	0.19	-0.01	-0.01	0.79	846	954
Behavioral risks (nonsexual) at baseline							
Number of times suspended or expelled	0.94	0.87	0.07	0.08	0.44	807	920
Ever skipped school	0.20	0.20	-0.01	-0.02	0.68	846	953
Proportion of substances used (vape, cigarettes, alcohol, marijuana, other illicit drugs)	0.10	0.11	-0.01	-0.03	0.59	756	866

Characteristic	MPC mean	Control mean	Difference	Effect size	*p*-value	MPC sample size	Control sample size
Contacted someone online to meet up in person	0.27	0.28	-0.01	-0.02	0.71	758	859
Survey timing							
Days between baseline and follow-up surveys	278.05	244.23	33.82	0.41	0.01	806	917
Presence of a summer between baseline and follow-up	0.62	0.48	0.13	0.27	0.19	846	954
Risk and protective factor baseline assessments[a]							
Knowledge about HIV/STIs	0.41	0.41	-0.01	-0.02	0.77	790	880
Knowledge about pregnancy	0.50	0.53	-0.02	-0.08	0.39	797	891
Knowledge about condoms	0.37	0.40	-0.03	-0.10	0.17	796	885
Knowledge about other forms of contraception	0.12	0.13	-0.01	-0.05	0.50	791	883
Belief that condoms can be pleasurable	0.13	0.15	-0.02	-0.05	0.37	777	878
Attitudes about condoms	0.69	0.66	0.02	0.09	0.07	796	889
Condom negotiation self-efficacy	0.83	0.83	0.00	0.01	0.93	795	895
Refusal self-efficacy	0.73	0.75	-0.01	-0.03	0.53	789	897
Sexual behavior baseline assessments[a]							
Ever any sex (vaginal, oral, or anal)	0.29	0.33	-0.03	-0.07	0.23	765	868
Any sex in the past three months	0.19	0.20	-0.01	-0.01	0.81	765	864
Times having any sex in the past three months	1.28	1.75	-0.47	-0.05	0.35	761	859
Count of vaginal sex partners in the past three months	0.83	1.22	-0.39	-0.09	0.12	738	826
Any sex without a condom in the past three months	0.13	0.14	-0.01	-0.02	0.70	762	854
Times having any sex without a condom in the past three months	0.70	1.13	-0.43	-0.05	0.30	762	854
Sex without birth control in the past three months	0.09	0.11	-0.02	-0.07	0.28	725	813
Times having sex without birth control in the past three months	0.25	0.68	-0.43	-0.10	0.04	725	813
MPC health goal baseline assessments[a]							
Ever pregnant	0.01	0.01	0.00	0.00	0.93	731	814
Ever had an STI	0.09	0.09	0.00	0.01	0.90	767	872

Sources: Baseline student survey, school rosters, and parent consent forms.

Note: Treatment and control means were model based, after adjusting for strata fixed effects. The *p*-values were based on an inferential test that adjusted the standard errors of the impact estimate for clustering. The baseline survey did not measure belief that sex may adversely affect future goals or condom use self-efficacy. In its impact analyses for these outcomes, the study team adjusted for proxies for these outcomes: belief that condoms can be pleasurable and condom negotiation self-efficacy, respectively. Missing data on the days between baseline and follow-up survey variable was predominantly due to individuals who completed a follow-up survey without completing the baseline.

[a] See Appendix A for details on how each outcome was operationalized.

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

Appendix C: Subgroup and sensitivity analyses

A. Subgroup analysis

The impacts of programs like MPC may vary across individuals or sites. For example, students who have previously had sex may respond more or less to certain programming, or may respond in different ways, than those who have never had sex. Features of implementation, context, or other idiosyncratic factors in a particular study site may magnify or dampen the program's impact compared to other locations. To explore these possibilities, the study team calculated separate impacts for MPC for several subgroups in the sample. Because of random assignment, subgroups such as gender were spread across both the treatment and control groups. As a result, even within a specific gender or location, the team had a valid experiment that could be used to estimate credible impacts on these groups, just as with the full sample.[29]

1. Sexually experienced and inexperienced youth

The study team first separated the sample of students into groups, depending on whether they reported at baseline having ever had sex. In the original effectiveness study of MPC, Jemmott et al. (1998) found that the program was particularly effective at improving behavioral outcomes for youth who reported at baseline being sexually active. In the current study, 34 percent of respondents reported prior instances of vaginal, anal, or oral sex at baseline.

Exhibit C.1 shows the results for the sexually experienced and inexperienced subgroups, alongside the benchmark results for comparison. The tables in this appendix present raw impact estimates for each sample; the Bayesian posterior probability of a favorable impact caused by MPC appears in parentheses. In both subsamples, the traditional statistical significance markers, based on *p*-values, are more or less consistent with the full sample—present among precursor outcomes and often suggesting a favorable impact of MPC, and absent among behavioral/health outcomes.[30] Though the reduction in the number of times respondents had sex in the past three months was not, as in the full sample, statistically significant in either subgroup, the point estimates suggest the reduction was driven by the subgroup who reported at baseline having sexual experience.

For behavioral outcomes, the Bayesian analysis suggests MPC had a favorable impact on most outcomes in the sexually experienced subgroup. Its impact on sexual behaviors among the sexually inexperienced subgroup was more nuanced. Whereas the Bayesian posterior probabilities suggest that more of the sexually inexperienced youth in the MPC group initiated sex compared to the control group, sexually inexperienced youth in the MPC group had sex without birth control less frequently and had lower rates of STIs. As noted above, none of the subgroup impact estimates on sexual behavior outcomes were statistically significant.

[29] For the third and fourth subgroup analyses described below, the study's regression specification omitted stratification variables because the clusters came from a limited number of strata. Otherwise, the specification remained the same as in the benchmark analysis throughout. This decision did not qualitatively affect the study's findings.

[30] The favorable full sample impacts on attitudes about condoms, condom use self-efficacy, and condom negotiation self-efficacy appear to have been driven by the sexually inexperienced subgroup.

Exhibit C.1. Impact estimates for sexually experienced and inexperienced subgroups

Outcome measure	Benchmark analysis	Sexually experienced	Sexually inexperienced
Risk and protective factor domain			
Knowledge about HIV/STIs	0.091*** (>99)	0.081*** (>99)	0.094*** (>99)
Knowledge about pregnancy	0.055*** (>99)	0.046 (96)	0.049* (99)
Knowledge about condoms	0.159*** (>99)	0.112*** (>99)	0.187*** (>99)
Knowledge about other forms of contraception	0.021* (98)	0.007 (70)	0.017 (93)
Belief that sex may adversely affect future goals	-0.032 (3)	-0.006 (48)	-0.028 (14)
Belief that condoms can be pleasurable	0.049* (>99)	0.051 (95)	0.041 (95)
Attitudes about condoms	0.055*** (>99)	0.045 (96)	0.044* (99)
Condom self-efficacy	0.058** (>99)	0.010 (68)	0.076** (>99)
Condom negotiation self-efficacy	0.043** (>99)	0.023 (81)	0.042* (99)
Refusal self-efficacy	0.041** (>99)	0.025 (83)	0.039 (98)
Sexual behavior domain			
Ever any sex (vaginal, oral, or anal)	0.030 (8)	n.a. (n.a.)	0.042 (12)
Any sex in past three months	-0.006 (63)	-0.045 (81)	0.015 (33)
Times having any sex in past three months	-1.300* (98)	-3.994 (96)	-0.184 (73)
Count of vaginal sex partners in the past three months	-0.185 (80)	-0.636 (81)	-0.009 (62)
Any sex without a condom in past three months	-0.007 (67)	-0.041 (77)	0.011 (39)
Times having any sex without a condom in past three months	-0.566 (87)	-1.322 (78)	-0.111 (66)
Sex without birth control in past three months	-0.022 (90)	-0.038 (73)	-0.007 (69)
Times having sex without birth control in past three months	-0.053 (59)	0.248 (47)	-0.163 (83)

Outcome measure	Benchmark analysis	Sexually experienced	Sexually inexperienced
MPC health goal domain			
Ever pregnant	0.008	0.022	-0.001
	(28)	(34)	(59)
Any STI	-0.007	0.012	-0.026
	(70)	(44)	(90)

Source: Baseline survey and six-month follow-up survey.

Note: Impact estimates presented in raw units (see Appendix A for descriptions of each outcome measure). Bayesian posterior probabilities presented in parentheses, representing the probability that MPC had a favorable effect on the 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.

*** Significantly different from zero at the 0.001 level, two-tailed test.

n.a. = not applicable.

2. Male and female youth

The study team also calculated impacts separately for male and female students.[31] A larger share of male students reported prior sexual initiation at baseline than among female students (45 percent versus 25 percent), though similar shares reported a previous STI diagnosis (9 percent) or pregnancy for themselves or a partner (2 percent).

Results among male and female students were substantively similar to the results among the sexually experienced and inexperienced subgroups, respectively, with the groups overlapping substantially (Exhibit C.2). Again, traditional *p*-values suggest no significant impacts on any behavioral or health outcomes for either of these subgroups. The Bayesian posterior probabilities suggest favorable impacts on nearly all behavioral outcomes for male students. They also suggest that more female students in the MPC group may have initiated sex and had recent sex, and that female students in the MPC group may have had sex less frequently. The behavioral impacts for both subgroups were small in magnitude, calculated on a smaller sample, and relatively noisy and nonsignificant.

Exhibit C.2. Impact estimates for female and male students

Outcome measure	Benchmark analysis	Subgroup analyses	
		Female	Male
Risk and protective factor domain			
Knowledge about HIV/STIs	0.091***	0.081***	0.110***
	(>99)	(>99)	(>99)
Knowledge about pregnancy	0.055***	0.041	0.072**
	(>99)	(98)	(>99)
Knowledge about condoms	0.159***	0.174***	0.142***
	(>99)	(>99)	(>99)
Knowledge about other forms of contraception	0.021*	0.014	0.036**
	(98)	(86)	(>99)

[31] Sample sizes were too small to examine the third gender category.

Outcome measure	Benchmark analysis	Subgroup analyses	
		Female	Male
Belief that sex may adversely affect future goals	-0.032	-0.030	-0.026
	(3)	(20)	(24)
Belief that condoms can be pleasurable	0.049*	0.065**	0.046
	(>99)	(>99)	(96)
Attitudes about condoms	0.055***	0.051**	0.068**
	(>99)	(>99)	(>99)
Condom self-efficacy	0.058**	0.052*	0.082**
	(>99)	(98)	(>99)
Condom negotiation self-efficacy	0.043**	0.039*	0.049
	(>99)	(98)	(98)
Refusal self-efficacy	0.041**	0.008	0.073**
	(>99)	(69)	(>99)
Sexual behavior domain			
Ever any sex (vaginal, oral, or anal)	0.030	0.047	0.017
	(8)	(7)	(33)
Any sex in past three months	-0.006	0.040	-0.039
	(63)	(11)	(88)
Times having any sex in past three months	-1.300*	-0.518	-2.183
	(98)	(83)	(95)
Count of vaginal sex partners in the past three months	-0.185	0.036	-0.503
	(80)	(23)	(80)
Any sex without a condom in past three months	-0.007	0.022	-0.031
	(67)	(27)	(84)
Times having any sex without a condom in past three months	-0.566	-0.193	-1.211
	(87)	(69)	(88)
Sex without birth control in past three months	-0.022	-0.004	-0.037
	(90)	(60)	(88)
Times having sex without birth control in past three months	-0.053	-0.021	-0.261
	(59)	(56)	(71)
MPC health goal domain			
Ever pregnant	0.008	0.012	0.010
	(28)	(27)	(39)
Any STI	-0.007	0.001	-0.007
	(70)	(54)	(63)

Source: Baseline survey and six-month follow-up survey.

Note: Impact estimates presented in raw units (see Appendix A for descriptions of each outcome measure). Bayesian posterior probabilities presented in parentheses, representing the probability that MPC had a favorable effect on the 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.

*** Significantly different from zero at the 0.001 level, two-tailed test.

3. Initial versus subsequent randomizations in each school

Because this study design re-randomized schools to a new experimental condition over multiple cohorts, the study team estimated separate impacts for schools based on their instance of random assignment. The first randomized cohort in each school included approximately 45 percent of the full sample. The second subgroup examined included the other 55 percent of the sample, as well as youth from the second or third randomized cohort in each school. This approach helped explore the possibility that cross-cohort interactions between students in the same school or increasing health educator experience (in Cohorts 2 and 3) might have affected observed impacts.

The timing of each city's participation in the trial means that the representation of different cities and schools across these subgroups is quite different than in the overall sample—in particular for Mobile and St. Louis schools. Though Mobile accounted for more than half of all participants in the overall sample, it represented 29 percent in the first randomization sample, and 84 percent in the subsequent randomization sample. Meanwhile, St. Louis represented 50 percent of the first randomization sample, but only 4 percent in the subsequent randomization sample. Detroit and Cincinnati schools were much more evenly balanced across the two subgroups, given that each school in those cities participated exactly once in Cohorts 2 and 3.

The impacts for risk and protective factor outcomes for these two groups were quite similar to the full sample benchmark results (Exhibit C.3). None of the behavioral outcomes showed significant results for the two subgroups. Bayesian posterior probabilities suggest that MPC may have had a favorable impact on several behavioral outcomes among youth in the first randomization subgroup, including the number of recent vaginal sex partners, likelihood of recent sex, and recent sex without a condom, but there was a potential increase in sexual initiation among this sample. The Bayesian posterior probabilities also suggest that MPC reduced the number of times that youth had sex among the second instance of randomization. As with the previous findings for sexual initiation and gender subgroups, these behavioral impacts were small in magnitude, relatively noisy, and not statistically significant.

Exhibit C.3. Impact estimates for first and subsequent rounds of random assignment within each school

Outcome measure	Benchmark analysis	Subgroup analyses	
		First instance of RA	Subsequent instances of RA
Risk and protective factor domain			
Knowledge about HIV/STIs	0.091*** (>99)	0.042 (94)	0.141*** (>99)
Knowledge about pregnancy	0.055*** (>99)	0.030 (89)	0.078*** (>99)
Knowledge about condoms	0.159*** (>99)	0.130*** (>99)	0.192*** (>99)
Knowledge about other forms of contraception	0.021* (98)	0.014 (87)	0.021 (87)
Belief that sex may adversely affect future goals	-0.032 (3)	-0.004 (52)	-0.048 (4)
Belief that condoms can be pleasurable	0.049* (>99)	0.006 (65)	0.076** (>99)

Outcome measure	Benchmark analysis	Subgroup analyses: First instance of RA	Subgroup analyses: Subsequent instances of RA
Attitudes about condoms	0.055***	0.022	0.073***
	(>99)	(85)	(>99)
Condom self-efficacy	0.058**	0.030	0.096***
	(>99)	(88)	(>99)
Condom negotiation self-efficacy	0.043**	0.015	0.054*
	(>99)	(78)	(99)
Refusal self-efficacy	0.041**	0.055*	0.025
	(>99)	(99)	(89)
Sexual behavior domain			
Ever any sex (vaginal, oral, or anal)	0.030	0.036	0.024
	(8)	(15)	(27)
Any sex in past three months	-0.006	-0.051	0.018
	(63)	(93)	(32)
Times having any sex in past three months	-1.300*	-0.622	-1.378
	(98)	(73)	(97)
Count of vaginal sex partners in the past three months	-0.185	-0.386	0.026
	(80)	(80)	(46)
Any sex without a condom in past three months	-0.007	-0.046	0.027
	(67)	(93)	(22)
Times having any sex without a condom in past three months	-0.566	-0.437	-0.358
	(87)	(71)	(68)
Sex without birth control in past three months	-0.022	-0.017	-0.007
	(90)	(74)	(65)
Times having sex without birth control in past three months	-0.053	0.218	-0.041
	(59)	(42)	(59)
MPC health goal domain			
Ever pregnant	0.008	0.018	0.017
	(28)	(25)	(21)
Any STI	-0.007	0.025	-0.016
	(70)	(40)	(78)

Source: Baseline survey and six-month follow-up survey.

Note: Impact estimates presented in raw units (see Appendix A for descriptions of each outcome measure). Bayesian posterior probabilities presented in parentheses, representing the probability that MPC had a favorable effect on the 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.

*** Significantly different from zero at the 0.001 level, two-tailed test.

RA = random assignment.

4. Geographic subgroups

Finally, the study team separated the analytic sample by city. Because Mobile contributed 18 out of the 31 total clusters for the sample, and St. Louis contributed the next largest number of clusters (n = 7), the

team estimated impacts separately for those two cities. As described in Section IV of the main report, MPC courses conducted in St. Louis schools displayed a heightened number of departures from the official curriculum and lower attendance rates. In addition, students in St. Louis responded to the follow-up survey at lower rates than students in other cities.

Results for these subgroups are shown in Exhibit C.4. Impacts on the precursor outcomes were generally quite favorable (except for attitudes about sex as a barrier to education completion and career) and statistically significant, but in St. Louis, the risk and protective factor impacts were less favorable (either smaller in magnitude or showing a negative impact estimate). The lack of statistically significant findings for the St. Louis subgroup was also likely caused by the low power of this analysis, based on just seven clusters.

Results for behavioral outcomes in Mobile were comparable to the full-sample results, with the nuance that the Bayesian posteriors for the Mobile subgroup suggest a favorable impact on STI incidence. For St. Louis, the Bayesian posteriors were frequently attenuated relative to the full sample (the probabilities of favorable/unfavorable effects were pulled toward 50 percent) for most outcomes, and the estimates were small, noisy, and not statistically significantly different from zero.

Exhibit C.4. Impact estimates for two geographic subgroups

Outcome measure	Benchmark analysis	Subgroup analyses	
		Mobile	St. Louis
Risk and protective factor domain			
Knowledge about HIV/STIs	0.091***	0.127***	-0.004
	(>99)	(>99)	(59)
Knowledge about pregnancy	0.055***	0.054**	-0.024
	(>99)	(>99)	(44)
Knowledge about condoms	0.159***	0.164***	0.058
	(>99)	(>99)	(83)
Knowledge about other forms of contraception	0.021*	0.011	-0.038
	(98)	(81)	(44)
Belief that sex may adversely affect future goals	-0.032	-0.035	0.001
	(3)	(12)	(62)
Belief that condoms can be pleasurable	0.049*	0.080**	-0.009
	(>99)	(>99)	(53)
Attitudes about condoms	0.055***	0.073***	-0.063
	(>99)	(>99)	(25)
Condom self-efficacy	0.058**	0.058*	0.007
	(>99)	(98)	(66)
Condom negotiation self-efficacy	0.043**	0.039	-0.042
	(>99)	(96)	(44)
Refusal self-efficacy	0.041**	0.029	0.035
	(>99)	(92)	(83)

Outcome measure	Benchmark analysis	Subgroup analyses	
		Mobile	St. Louis
Sexual behavior domain			
Ever any sex (vaginal, oral, or anal)	0.030	0.017	0.010
	(8)	(34)	(49)
Any sex in past three months	-0.006	-0.003	-0.109
	(63)	(58)	(76)
Times having any sex in past three months	-1.300*	-1.814	-0.716
	(98)	(94)	(84)
Count of vaginal sex partners in the past three months	-0.185	-0.436	0.062
	(80)	(83)	(49)
Any sex without a condom in past three months	-0.007	0.006	-0.128
	(67)	(47)	(76)
Times having any sex without a condom in past three months	-0.566	-0.843	-0.299
	(87)	(90)	(75)
Sex without birth control in past three months	-0.022	-0.020	-0.027
	(90)	(79)	(70)
Times having sex without birth control in past three months	-0.053	-0.307	0.047
	(59)	(78)	(51)
MPC health goal domain			
Ever pregnant	0.008	0.015	0.028
	(28)	(26)	(32)
Any STI	-0.007	-0.029	-0.002
	(70)	(87)	(57)

Source: Baseline survey and six-month follow-up survey.

Note: Impact estimates presented in raw units (see Appendix A for descriptions of each outcome measure). Bayesian posterior probabilities presented in parentheses, representing the probability that MPC had a favorable effect on the 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.

*** Significantly different from zero at the 0.001 level, two-tailed test.

B. Sensitivity tests

As described in Section III in the main report, in addition to its benchmark analysis, the study team estimated impacts using seven other approaches to understand the robustness of its findings across different modeling and data preparation decisions.

1. **Benchmark:** Analysis of consistent survey responses with observed data for follow-up measures. The study team made statistical adjustments for a large baseline covariate set (with dummy variable indicators for missing baseline data and baseline variables imputed to a constant), adjustment for two variables measuring period of time between baseline and follow-up surveys, and strata fixed effects.
2. **Raw data:** Identical model specification to benchmark, but instead of consistent survey data, the study team conducted the analysis on raw survey responses (including inconsistent survey responses).

3. **Complete case:** Analysis of raw survey responses, but the study team dropped observations with missing responses for any baseline covariate.
4. **Multiple imputation:** Multiple imputation of all missing baseline and follow-up data. The study team made statistical adjustment for a large baseline covariate set, adjustment for two variables measuring period of time between baseline and follow-up surveys, and strata fixed effects.
5. **No strata:** Identical model specification as benchmark, except the study team excluded strata fixed effects.
6. **Parsimonious:** Analysis of consistent survey responses that included only strata fixed effects.
7. **No adjustment for survey timing:** Identical model specification as benchmark, except the study team excluded the two survey timing variables.
8. **Fully Bayesian approach:** Calculation of Bayesian posterior probabilities was based on a Bayesian meta-regression of MPC impact estimates.

All analytic approaches clustered standard errors to account for the non-independence of student survey responses within each of the 31 units of assignment.

Exhibit C.5 summarizes the impact estimates for the benchmark and these seven alternative approaches. The summary of findings below focuses on the first six columns of sensitivity analyses, because the impact estimates and *p*-values for the Fully Bayesian approach are identical to the Benchmark approach—only the posterior probabilities differ. In the Fully Bayesian approach, nearly all posterior probabilities have been pulled slightly towards a 50 percent probability, aligning with expectations that impact estimates in a given domain are correlated. In this specification, the probability that MPC has a favorable effect on one outcome in the domain acknowledges the likelihood that other outcomes in the domain also have favorable effects.

The direction, magnitude, and significance of impacts observed in the other sensitivity approaches were quite similar to those observed for the benchmark approach for nearly all outcomes examined. This finding suggests that this study's impact estimates were not driven by methodological choices. There were a few departures—in both the more and less significant directions—from full consistency on which impact estimates meet traditional *p*-value significance thresholds. Of the 120 impact estimates presented for the six main sensitivity analyses examined, only 10 had different significance findings from the benchmark analysis, and seven had impact estimates that differed in direction. Importantly, the change in direction occurred only among nonsignificant benchmark findings. Therefore, given that the overwhelming majority of findings across the sensitivity analyses shared the same direction and statistical significance, this finding suggests that the benchmark analytic approach is a defensible and robust manner to summarize the effect of MPC across all outcomes examined.

Exhibit C.5. Impact estimates under six sensitivity analyses

Outcome measure	Benchmark analysis	Sensitivity analyses						
		Raw data	Complete case	Multiple imputation	No strata	Parsimonious	No adjustment for survey timing	Fully Bayesian
Risk and protective factor domain								
Knowledge about HIV/STIs	0.091*** (>99)	0.090*** (>99)	0.083*** (>99)	0.087*** (>99)	0.094*** (>99)	0.090*** (>99)	0.096*** (>99)	0.091*** (>99)
Knowledge about pregnancy	0.055*** (>99)	0.053*** (>99)	0.047** (>99)	0.055** (>99)	0.064*** (>99)	0.056* (98)	0.060*** (>99)	0.055*** (>99)
Knowledge about condoms	0.159*** (>99)	0.158*** (>99)	0.160*** (>99)	0.159*** (>99)	0.167*** (>99)	0.149*** (>99)	0.162*** (>99)	0.159*** (>99)
Knowledge about other forms of contraception	0.021* (98)	0.020* (98)	0.012 (88)	0.016 (95)	0.024 (97)	0.020 (89)	0.027** (>99)	0.021* (98)
Belief that sex may adversely affect future goals	-0.032 (3)	-0.032 (4)	-0.027 (9)	-0.031 (6)	-0.035 (3)	-0.038 (5)	-0.032 (5)	-0.032 (8)
Belief that condoms can be pleasurable	0.049* (>99)	0.049* (99)	0.056** (>99)	0.043* (98)	0.051** (>99)	0.057** (>99)	0.051** (>99)	0.049* (99)
Attitudes about condoms	0.055*** (>99)	0.053*** (>99)	0.051*** (>99)	0.057*** (>99)	0.052*** (>99)	0.065*** (>99)	0.055*** (>99)	0.055*** (>99)
Condom self-efficacy	0.058** (>99)	0.057** (>99)	0.072*** (>99)	0.061** (>99)	0.059** (>99)	0.071*** (>99)	0.063** (>99)	0.058** (>99)
Condom negotiation self-efficacy	0.043** (>99)	0.042** (>99)	0.046** (>99)	0.043* (99)	0.043** (>99)	0.067*** (>99)	0.051*** (>99)	0.043** (>99)
Refusal self-efficacy	0.041** (>99)	0.041** (>99)	0.033 (97)	0.041* (99)	0.033* (99)	0.036* (98)	0.045** (>99)	0.041** (99)

Outcome measure	Benchmark analysis	Sensitivity analyses						
		Raw data	Complete case	Multiple imputation	No strata	Parsimonious	No adjustment for survey timing	Fully Bayesian
Sexual behavior domain								
Ever any sex (vaginal, oral, or anal)	0.030 (8)	0.015 (28)	0.008 (40)	0.029 (12)	0.024 (11)	0.014 (33)	0.039* (3)	0.030 (16)
Any sex in past three months	-0.006 (63)	-0.008 (67)	-0.006 (62)	0.001 (50)	-0.016 (80)	-0.008 (67)	0.002 (50)	-0.006 (61)
Times having any sex in past three months	-1.300* (98)	-1.139 (95)	-2.172** (>99)	-1.084 (94)	-1.286* (98)	-1.221 (93)	-1.144 (97)	-1.300* (87)
Count of vaginal sex partners in the past three months	-0.185 (80)	-0.166 (81)	-0.296 (85)	-0.162 (81)	-0.189 (83)	-0.146 (79)	-0.158 (78)	-0.185 (67)
Any sex without a condom in past three months	-0.007 (67)	-0.010 (70)	-0.010 (69)	-0.012 (71)	-0.005 (64)	-0.006 (62)	0.001 (50)	-0.007 (62)
Times having any sex without a condom in past three months	-0.566 (87)	-0.355 (76)	-1.044 (93)	-0.552 (90)	-0.532 (86)	-0.567 (87)	-0.433 (81)	-0.566 (71)
Sex without birth control in past three months	-0.022 (90)	-0.020 (84)	-0.019 (75)	-0.013 (74)	-0.012 (78)	-0.013 (75)	-0.014 (80)	-0.022 (79)
Times having sex without birth control in past three months	-0.053 (59)	0.084 (41)	-0.310 (82)	-0.010 (54)	0.089 (41)	0.015 (51)	0.005 (52)	-0.053 (46)
MPC health goal domain								
Ever pregnant	0.008 (28)	0.013 (16)	0.010 (26)	0.007 (31)	0.011 (18)	0.016 (18)	0.009 (23)	0.008 (32)
Any STI	-0.007 (70)	-0.007 (69)	-0.004 (62)	-0.007 (65)	-0.005 (63)	-0.008 (73)	-0.009 (75)	-0.007 (68)

Source: Baseline survey and six-month follow-up survey.

Note: Impact estimates presented in raw units (see Appendix A for descriptions of each outcome measure). Bayesian posterior probabilities presented in parentheses, representing the probability that MPC had a favorable effect on the 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.

*** Significantly different from zero at the 0.001 level, two-tailed test.

OASH | Office of Population Affairs

Office of Population Affairs | Website: https://www.hhs.gov/opa/ Email: opa@hhs.gov Phone: (240) 453-2846