Contract Number:
HHSP23320082911YC

Mathematica Reference Number:
06549.070

Submitted to:
Office of Adolescent Health
Division of Policy, Planning & Communication
Department of Health & Human Services
1101 Wootton Parkway, Suite 700
Rockville, MD 20852

Project Officer: Amy Farb

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Alan Hershey

Evaluation of Adolescent Pregnancy Prevention Approaches: Design of the Impact Study

Final Report

October 2012

Primary Authors:
Kimberly Smith, Mathematica
Silvie Colman, Mathematica

With Support From:

Mathematica:
Christopher Trenholm
Alan Hershey
Brian Goesling
Anastasia Erbe
Caitlin Davis
Brice Overcash

Child Trends:
Kristine Andrews
Amanda Berger
Lori Ann Delale-OConnor
Mindy Scott

MATHEMATICA
Policy Research

THIS PAGE IS INTENTIONALLY BLANK

CONTENTS

TABLES

FIGURES

THIS PAGE IS INTENTIONALLY BLANK

I. OVERVIEW OF THE IMPACT EVALUATION

Adolescent sexual activity and its consequences continue to be important policy concerns in the United States. Nationwide, nearly half of all high school students report having had sex and one-fifth report having had four or more partners by the time they graduate (Centers for Disease Control and Prevention [CDC] 2010a). In 2009, almost 40 percent of sexually active high school students did not use a condom during their last sexual intercourse, and 12 percent did not use any method of contraception (CDC 2010a). Such sexual risk behaviors place adolescents at risk for unintended pregnancy and sexually transmitted diseases (STDs). In 2009, over 400,000 infants were born to 15- to 19-year-olds, and nearly half of the 19 million new STD cases each year are among youth in this age group (CDC 2010b). These and related social and economic consequences of teen sexual behaviors can be detrimental to teens, their families, their communities, and society as a whole.

The Evaluation of Adolescent Pregnancy Prevention Approaches (PPA) is a response to persistent concerns about the consequences of teen sexual activity. The PPA evaluation is being undertaken to expand available evidence on effective ways to prevent and reduce pregnancy and related sexual risk behaviors among teens in the United States. This eight-year (2008–2016) evaluation is being conducted by Mathematica Policy Research and its partners, Child Trends and Twin Peaks Partners LLC, under contract to the Office of Adolescent Health (OAH) in the Office of the Assistant Secretary for Health of the U.S. Department of Health and Human Services (HHS). It will document and rigorously test promising pregnancy prevention approaches in seven sites across the United States, each of which will implement a different program. The evaluation has two components: (1) an in-depth implementation analysis of the seven selected programs and (2) a rigorously designed impact study of each program. The implementation analysis will document and examine the development, implementation, and operations of the selected programs. The impact studies will use experimental designs and longitudinal survey data in all sites, and focus on assessing the effectiveness of each selected program on its own, compared to a control group in the same site.

The purpose of this report is to describe the impact study component of the PPA evaluation. This chapter provides an overview of the impact evaluation, including the policy context and goals for the evaluation (Section A), the selection and key features of the program models and sites included in the evaluation (Section B), the impact study designs in the seven sites (Section C), and our approach to estimating impacts and reporting results (Section D). Chapters II through VIII focus on the seven individual sites, providing more detailed information on the program model, evaluation design, and analytic approach for each site.

A. Background and Goals for the Evaluation

Despite a one-third decline since the early 1990s, the United States still has the highest rates of teen pregnancy and birth among industrialized countries. The U.S. teen birth rate is roughly one and a half times higher than the rate in the United Kingdom, which has the highest teen birth rate in Europe (United Nations 2010). Approximately three-quarters of a million women below the age of 20 in the United States became pregnant in 2006 (Kost et al. 2010). In 2009, teen births accounted for 10 percent of all births and one-fifth of all nonmarital births in the United States (Martin et al. 2011).

The economic and social costs of teen births are substantial. A majority of teen mothers live in poverty at the time of their child's birth and are less likely to finish high school and achieve economic self-sufficiency than adult mothers. Children born to teen mothers are at greater risk of

adverse health outcomes and are more likely to live in single-parent households, to have low academic achievement, and to give birth themselves in their teens (Hoffman 2008).

Recent data suggest that changes in teen sexual risk behaviors that accompanied the post-1991 decline in teen birth rates have stalled, and concerns about consequences persist. While the proportion of high school students who have ever had sexual intercourse declined from 54 percent in 1991 to 46 percent in 2009, that proportion remained relatively constant between 2003 and 2009 (CDC 2010a). In addition, the magnitude of the decline since the early 1990s varies across racial and ethnic lines, with some high school populations—such as Hispanic female students—showing increased rates of sexual activity over this period (CDC 2010a). In 2009, more than one-third of high school students were sexually active; fewer than two-thirds of the sexually active students had used a condom at last intercourse and only 9 percent had used dual methods of contraception (CDC 2010a). In addition to the risk of teen pregnancy, teen sexual activity also brings increased risks of STDs. Compared to older adults, sexually active teens are at greater risk of acquiring STDs due to both biological and behavioral factors. Teens between the ages of 15 and 19 have the highest reported case rates for Chlamydia and gonorrhea of any age group (CDC 2010b). Among sexually active teenage girls, STD prevalence is estimated to be roughly 40 percent (Hampton 2008). STDs have been linked to infertility, pregnancy complications, cervical cancer, increased HIV risk, and numerous other health problems.

Mounting concerns over the risks of teen sexual activity have spurred the creation of a broad range of program approaches to reducing high rates of teen pregnancy and STDs among U.S. adolescents. The range of existing program strategies reflects the complexity of the issue and the varying needs of diverse target populations, as well as differing stakeholder opinions about the most appropriate strategy. Abstinence education programs focus on delaying sexual initiation, and many of these programs focus specifically on delaying sexual activity until marriage. By contrast, comprehensive sex education approaches generally include information on (or otherwise promote) both the benefits of abstinence and risk mitigation through condom and contraceptive use for sexually active adolescents. Youth development approaches often combine elements of either abstinence or comprehensive sex education approaches with broader services, such as mentoring, health services, or case management. Some of these varied program models target widely defined populations in diverse settings, while others were developed to meet the needs of specific high-risk populations. Many adolescent pregnancy prevention programs continue to be based in schools, but increasingly programs are also delivered in community, clinic, and home settings.

To assess the evidence on the effectiveness of individual programs across the range of approaches, HHS in 2009 launched the Teen Pregnancy Prevention Research Evidence Review (PPRER)—an ongoing, systematic review of the evidence base on approaches to preventing pregnancy, STDs, and associated sexual risk behaviors. From the 200 program impact studies identified and assessed in a first phase of PPRER (2009–2010), the review team identified 28 programs that met HHS criteria for effective evidence-based programs (Goesling and Trenholm 2010). In addition, the review team identified for HHS several weaknesses and gaps in the existing teen pregnancy prevention research literature. Although the literature includes experimental studies, the review team found significant variation in the quality and rigor of both random assignment and quasi-experimental studies, with only half of the 200 studies reviewed meeting the HHS-approved standards for research quality. In addition, limited evidence was found on several high-priority program settings, approaches, and target populations. These include in-school and rural programs, and programs for several high-risk populations, including high-school-age youth, pregnant and parenting adolescents, Latinos, and youth living in foster care.

HHS recently launched two major new federal programs to support the implementation of evidence-based programs to prevent teen pregnancy and to further expand the evidence base on effective program approaches. The first is the Teen Pregnancy Prevention (TPP) program, established in December 2009 by the Consolidated Appropriations Act of 2010. Administered by the newly created Office of Adolescent Health in HHS, the TPP program (2010–2015) provides competitive grants to local organizations, school districts, and state agencies to implement evidence-based and promising new teen pregnancy prevention programs. Of the funds available, $75 million is for replication of the programs shown to be effective through rigorous evaluation (Tier 1 programs) and $25 million is for research and demonstration programs to develop and test additional models and innovative new programs (Tier 2 programs). The second major federal program is the Personal Responsibility Education Program (PREP), authorized under the 2010 Patient Protection and Affordable Care Act and administered by the Administration for Children and Families (ACF). PREP provides grants to states, tribes, and tribal communities to replicate, or substantially incorporate elements of, programs that have been shown to be effective at delaying sexual initiation, reducing pregnancy, and increasing contraceptive use among sexually active youth. The selected programs must educate youth on both abstinence and contraception. They must also incorporate lessons on "adulthood" preparation topics, such as healthy relationships and parent-child communications. The majority of PREP's $75 million in annual funding (2010–2014) is reserved for the implementation of evidence-based teen pregnancy prevention programs, largely through formula grants to states and territories. However, PREP provides approximately $10 million annually for competitive grants to public and private entities to develop and test innovative strategies—Personal Responsibility Education Innovative Strategies (PREIS)—to reduce teen pregnancy and repeat pregnancy among vulnerable youth populations, such as youth who are homeless or in foster care.

The PPA evaluation is the first of several current large-scale federal evaluation efforts aimed at expanding the evidence base on pregnancy prevention approaches. In contrast to other federal evaluation efforts, which focus largely on replications of evidence-based programs, the PPA evaluation will document and test the effectiveness of untested and innovative programs that have not been rigorously evaluated. The seven programs selected for the evaluation offer a range of implementation settings, program strategies, and target populations, reflecting the diversity of the current program landscape. For this reason, the impact analysis of each program will serve as an independent study of the effectiveness of a single program, and results will not be compared or combined across programs. As a complement to the impact study in each site, the PPA evaluation will also include an implementation analysis examining the context and delivery of each program and the contrast between the services received by program participants and the counterfactual condition (what they would have received in the program's absence). The implementation study will provide a basis for interpreting estimates of program impacts and understanding how any programs that may be found effective might be replicated.

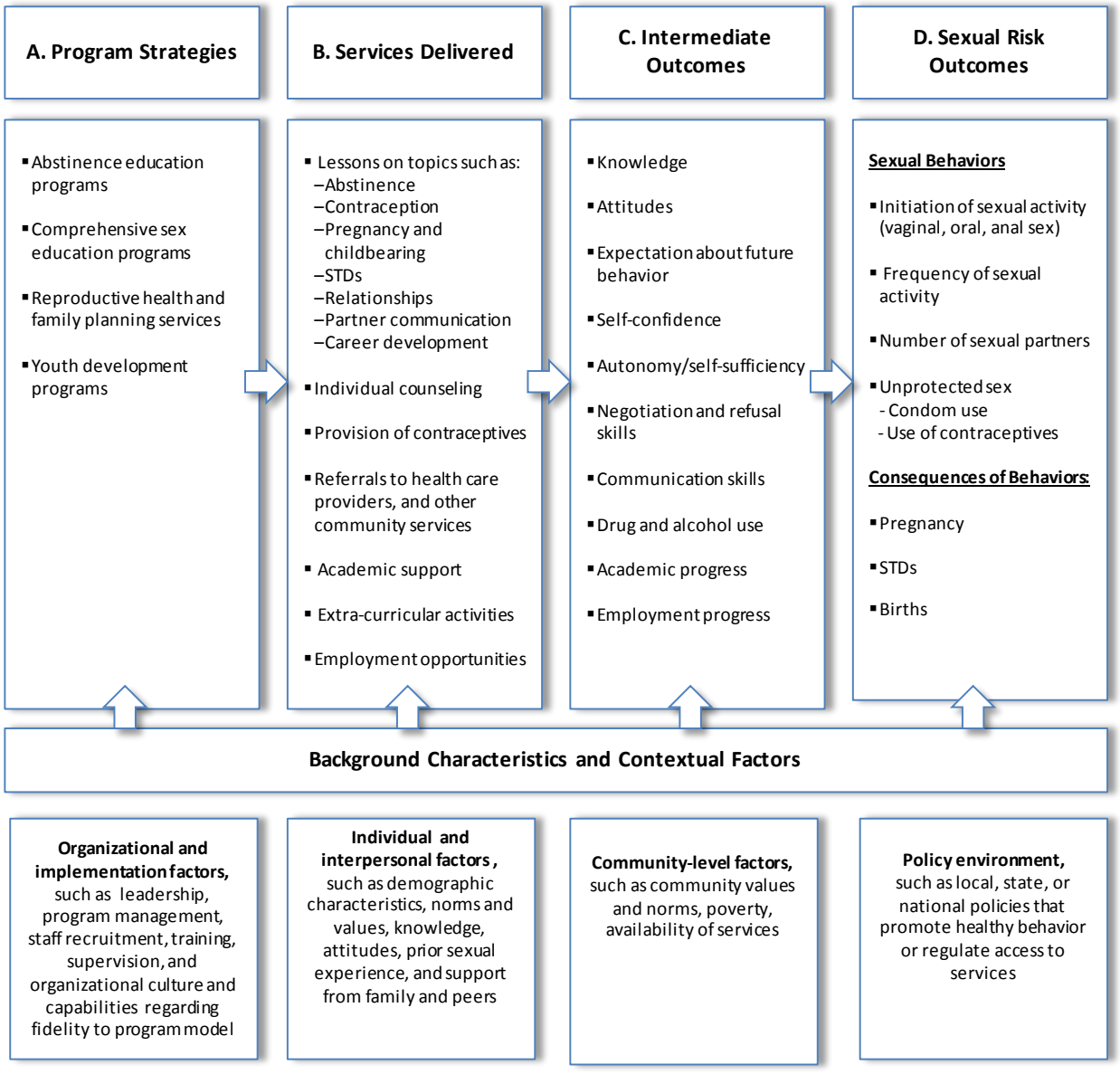
In each of the seven evaluation sites, the PPA evaluation team is working closely with the local organization(s) implementing the programs to ensure successful execution of the study design. As detailed in subsequent chapters, six of the seven programs in the PPA evaluation are being implemented by local organizations with TPP Tier 2 or PREIS grant funding. In these six sites, the PPA evaluation team is collaborating closely with the independent local evaluators, funded as part of each TPP/PREIS grant, on the design and implementation of the PPA impact and implementation studies. In carrying out this collaboration, the PPA evaluation team will have primary responsibility for all aspects of the PPA study design, including lead authorship of the reports completed as part of the PPA project. Local evaluators may also conduct their own analysis and reporting beyond the

scope of the PPA evaluation, drawing on their TPP/PREIS grant funding, with the schedule of that reporting determined by their funding agency (OAH or ACF).

B. Selection and Description of the Evaluation Sites

A conceptual framework has guided decisions on the design of the PPA evaluation (Figure I.1). The framework illustrates the process by which teen pregnancy prevention approaches might affect sexual risk behaviors and related outcomes. Pregnancy prevention programs take place in diverse and dynamic environments. The implementing organization and system, individual characteristics of youth, familial and peer support systems, community norms and resources, and policy-related factors all influence the development and selection of teen pregnancy prevention approaches, the specific services provided under each program, and how services are delivered. Individual-, community-, and policy-level factors also influence prior sexual behaviors and the availability of and access to existing services, as well as participation in teen pregnancy prevention programs and subsequent behavioral choices and outcomes.

Figure I.1. Conceptual Framework Guiding the PPA Evaluation



Pregnancy prevention programs (Column A of Figure I.1) aim to change, enhance, or supplement existing services and support that youth receive in their communities. Programs often provide information on health, relationship, and sex education topics, and may also address related social and behavioral issues, including life skills development, mental health, access to contraceptives and other health services, academic performance, and employment opportunities (Column B). Receipt of these additional services is hypothesized to have favorable impacts on intermediate outcomes (Column C) that may serve as mediators of sexual risk behaviors and their consequences. For example, youth in pregnancy prevention programs might increase their knowledge about sexual behavior risks, develop more positive views on abstinence and delaying pregnancy, improve their communication skills and relationship quality, reduce consumption of drugs and alcohol, or improve their academic performance. Through these and other changes, programs aim to affect sexual risk behaviors and, in turn, the incidence of pregnancy and STDs.

The PPA evaluation will focus largely on measuring the impacts of selected pregnancy prevention programs on the sexual risk outcomes shown in Column D of Figure I.1. However, as part of the impact analysis, we will also measure several intermediate outcomes that the programs under study are hypothesized to affect. More specifically, the impact evaluation will address the following three broad questions:

- **What impacts do the selected programs have on the key sexual risk outcomes targeted by the evaluation?** The particular outcomes tested will vary by site, but will be limited to the key measures of sexual risk behavior and its consequences in Column D of Figure I.1.
- **For which youth are the selected programs effective?** Understanding how the effectiveness of the teen pregnancy prevention approaches varies by characteristics of participating youth can suggest ways to target the approaches. This exploratory analysis will focus on a limited number of key subgroups. The subgroups of interest will vary across program models and target populations, but may include groups of youth defined by sex, baseline sexual experience, age, or race/ethnicity.
- **What are the pathways or mechanisms through which the programs work?** What impacts do programs have on possible mediators of key sexual behavior outcomes (such as those shown in Column C of Figure I.1)? For programs that successfully reduce sexual risk behavior, what role might these changes in mediators play? To explore these questions, we will estimate program impacts on key mediating factors identified in the site-specific program logic models.

These questions will be examined through comparisons of all youth in randomly created program and control groups, but youth who have access to the programs will participate at different levels. Some youth may participate in all sessions or activities, some may participate in a limited number, and some may not participate at all. As an exploratory analysis, we will examine differences in youth participation rates and variation in impacts by participation levels.

1. Selection of Sites

In choosing sites for the PPA evaluation, we sought up to eight sites where promising intervention approaches could be rigorously evaluated and where evaluation results would add to current knowledge about program effectiveness. The selection of evaluation sites involved two interrelated components: (1) identifying promising program approaches and (2) identifying program

sites that would allow for a high quality evaluation of the program. In collaboration with HHS, the evaluation team developed criteria for assessing the fit of program models and sites with the goals of the PPA evaluation.

To be considered for the study, a program approach had to meet three criteria: (1) it had to be policy-relevant and or offer a chance to fill gaps in the existing research literature on new program approaches, underserved populations, and/or program settings; (2) it had to be based on an evidence-based logic model linking the services delivered to key behavioral outcomes of central interest in the PPA evaluation; and (3) it had to be of sufficient intensity and duration that it is plausible to expect the program could affect key behaviors.

Assessment of potential program sites also involved determining the feasibility of successful implementation of a rigorous experiment. To qualify as an evaluation site, it had to be shown that the site (1) presented conditions suitable for implementing random assignment, (2) could achieve the sample size requirements for a rigorous study, (3) offered conditions in which contamination between the treatment and control groups could be minimized, and (4) would provide a meaningful contrast between the treatment and control conditions. In addition, the implementing organization(s) had to be able to deliver the program with sufficient fidelity to ensure a valid test of its impact. The final criterion used for site selection was the contribution of the program and site to the whole evaluation and its aim of including a group of programs and local sites that would maximize the overall knowledge generated by the evaluation.

Using these criteria, PPA evaluation sites were selected in two phases. During the first phase (2008–2010), the evaluation team cast a wide net to identify candidate program approaches and sites. To identify candidate programs, the evaluation team first reviewed existing research, conducted interviews with a wide range of stakeholders (including program developers, government agencies, researchers, professional organizations, and advocacy organizations), and visited programs currently in operation. The evaluation team then held discussions with program developers to gain additional information on promising interventions, the conditions needed for implementation fidelity, and planned replications or expansions that might present an opportunity for evaluation. Efforts to identify potential sites during this first phase focused on large school districts across the United States, and involved discussions with a large number of schools as well as statewide pregnancy prevention coalitions. Following in-depth discussions with several school districts in various states, one evaluation site was selected—Chicago Public Schools, which was implementing the *HealthTeacher* curriculum. In addition, the *Teen PEP* program, implemented by the program developer—Princeton Center for Leadership Training—in several New Jersey high schools, was selected during this first phase of site recruitment.

The second phase of the site recruitment effort focused primarily on identifying sites among the TPP Tier 2 and PREIS grantees. In December 2010, the evaluation team and HHS identified 19 grantees as potential sites. Following discussions with the OAH and ACF project officers overseeing the 19 grantees, the evaluation team held conference calls with candidate grantees and their independent evaluators. These calls were used to assess the feasibility of including grantees in the PPA evaluation, as well as to assess the relative benefits of selecting the grantee as a PPA site versus having the grantee simply continue with its own local evaluation required as a condition of grant award. Following further in-depth discussions between the evaluation team and a smaller set of grantees, six TPP/PREIS grantees were selected for inclusion in the study. These grantees are implementing the following six programs: *AIM 4 Teen Moms*, *Gender Matters (GEN.M), POWER Through Choices* 2010 (PTC 2010) *Teen Options to Prevent Pregnancy (T.O.P.P.)*, *Teen PEP* (implemented in

North Carolina), and *WAIT Training*. (The *Teen PEP*/North Carolina site will be combined with the *Teen PEP*/New Jersey site in order to form a sample of schools large enough for a rigorous evaluation of the program. Given this pooling, we discuss the evaluation of this program—as we do with other programs—in the context of a single site.)

2. Key Features of Program Models

The seven selected programs represent a range of program approaches, target populations, and evaluation settings (Table I.1). Two of the programs—*Aim 4 Teen Moms* and *T.O.P.P.*—focus on delaying repeat pregnancy among low-income pregnant and parenting teens, but use different approaches to achieve their program goals. *Aim 4 Teen Moms* uses an expanded and adapted version of an evidence-based youth development program (*Project Aim*); in contrast, *T.O.P.P.* is a clinic-based program that provides telephone-based care coordination and mobile contraceptive services. One of the selected programs, *WAIT Training*, is a school-based abstinence-until-marriage curriculum delivered in middle schools. Three programs represent varied comprehensive sex education approaches, each targeting and tailored to very different populations. These include: (1) an enhanced *HealthTeacher* curriculum implemented in middle schools; (2) *POWER Through Choices*, a sex education curriculum developed for foster care youth; and (3) *Teen PEP,* a high school-based peer-led sex education curriculum. Finally, *GEN.M* is a multicomponent youth development program that recruits youth applying for a summer employment program and is delivered the week after the employment program ends.

Other program features, and sources of program variation, include the following (Table I.1):

- **Program location.** The programs serve youth in a mix of urban and rural areas across the United States, including in Los Angeles (*Aim 4 Teen Moms*), Austin (*GEN.M*), and Chicago (*HealthTeacher*), and, in the case of other programs, in varied settings in California, Illinois, Maryland, New Jersey, North Carolina, Ohio, and Oklahoma.
- **Age of study population.** *HealthTeacher* and *WAIT Training* target 7th grade middle school students and *Teen PEP* serves 9th grade high school students. The four other programs target a broader age range, including low-income mothers ages 15 to 19 (*Aim 4 Teen Moms*) and 10 to 19 (*T.O.P.P.*), youth ages 14 to 16 (*GEN.M*), and foster care youth ages 14 to 18 (*POWER Through Choices*).
- **Program duration and intensity.** The duration and intensity of services vary widely, from five 90-minute sessions (*Teen PEP*) to monthly one-hour phone calls and periodic home visits delivered over an 18-month period (*T.O.P.P.).*
- **Delivery setting.** *HealthTeacher*, *Teen PEP*, and *WAIT Training* serve youth in school settings. *POWER Through Choices* delivers services in group foster homes. The other three programs serve youth in multiple settings, including in participants' homes and clinics (*Aim 4 Teen Moms*), in various community sites (*GEN.M*), and via telephone and in-person visits (*T.O.P.P.*).
- **Staff delivering the program.** The type of staff or volunteers delivering the program also varies across programs and delivery settings. The three school-based programs use different types of facilitators, all trained by program staff: teachers for *WAIT Training*, a mix of teachers and counselors for *HealthTeacher*, and student facilitators for *Teen PEP*. Other programs use facilitators hired and trained by the grantee (*Aim 4 Teen Moms*, *GEN.M*, and *POWER Through Choices*). *T.O.P.P.* is the only program that uses trained nurses to deliver program content.

Table I.1. Key Features of Programs Evaluated, by Site

Program Model (Implementing Organization)	Program Description	Study Location	Study Target Population	Program Duration and Intensity	Delivery Setting	Staff Delivering Program
AIM 4 Teen Moms (Children's Hospital Los Angeles)	Individual- and group-based youth development program that encourages parenting teens to delay repeat pregnancies	Los Angeles, CA	Low-income mothers ages 15 to 19 with a child between the ages of 1 and 6 months	Seven 60-minute sessions and two 90-minute sessions delivered over 12 weeks	Participants' homes and community spaces	Trained advisors
Gender Matters (GEN.M) (EngenderHealth; Austin, TX)	Multicomponent youth development program that aims to challenge traditional views on gender and incorporates comprehensive sexual education	Austin, TX	Youth ages 14 to 16 enrolled in the Travis County Summer Youth Employment Program	Five 4-hour sessions delivered over five consecutive days, plus a text messaging campaign and community event	Various community settings	Trained facilitators
HealthTeacher™ Curriculum, Enhanced (Chicago Public Schools)	Internet-accessible, school-based comprehensive sex education curriculum	Chicago, IL	7th grade students	Twelve 45- to 90-minute sessions delivered over a semester	Classroom	Trained teachers
POWER Through Choices (Oklahoma Institute for Child Advocacy)	Group-based sex education program targeting foster care youth	CA, IL, MD, and OK	Foster care youth ages 13 to 18	Ten 1.5-hour sessions over 4–12 weeks	Group foster homes	Trained facilitators
Teen Options to Prevent Pregnancy (T.O.P.P.) (OhioHealth Research and Innovation Institute)	Clinic-based program providing telephone-based care coordination and mobile contraceptive services	Central OH	Low-income OhioHealth patients ages 10 to 19 who are currently pregnant or recently delivered	Monthly 1-hour telephone calls and periodic home visits over 18 months	Telephone contact, home visits, mobile unit visits	Trained nurse educators
Teen PEP (Princeton Center for Leadership Training)	School-based peer-led comprehensive sex education workshops	NC and NJ	9th grade students	Five 1.5-hour workshops delivered over a semester (NC) or school year (NJ)	Classroom/ auditorium/gym	Trained student facilitators
WAIT Training (Live the Life)	School-based abstinence until marriage curriculum	Central and northern FL, and southern GA	7th grade students	Sixteen 1-hour sessions—8 hours in each of two consecutive school years	Classroom	Trained teachers

C. Overview of Impact Study Designs

As suggested by the variety of evaluation sites, the specific study designs must also diverge to some extent. Research questions and the outcomes of greatest interest build on the overall evaluation goals and design, but reflect the specific program approaches and logic models of each site. The populations they target, and thus the evaluation samples, will differ in characteristics and size. Since programs differ in their duration, and the trajectory of expected sexual activity varies with the age of the target populations, the schedules of data collection to support impact analysis will also vary by site.

1. Research Questions and Key Outcomes

In each site, the impact analysis will address the following primary research question: "Was the program successful at reducing teen pregnancy, STDs, or associated sexual risk behaviors?" As a formal test of this question, we will conduct a confirmatory analysis in each site focusing on only a small number of sexual risk outcomes that are central to the program model and can be rigorously evaluated. The "confirmatory outcomes" that are the focus of this analysis will vary by site and be tailored to the program's theory of change and the evaluation setting. While all of the programs being tested ultimately aim to reduce teen pregnancies and related teen risk behaviors, their specific objectives within the study period vary with the program model and target population. Some programs being tested focus more on promoting abstinence, others on sexual risk behaviors and use of contraceptives. Several programs (*HealthTeacher*, *Teen PEP*, and *WAIT Training*) target young populations in which pregnancy can be expected to be a rare outcome during the study period. In these sites, a confirmatory impact analysis focused on the outcome of pregnancy would not have sufficient statistical power to measure program impacts within the study period, so instead the confirmatory analysis will focus on sexual risks closely tied to pregnancy, such as initiation of sexual (vaginal) intercourse and/or prevalence of unprotected intercourse.

Beyond the narrow confirmatory analysis, we will conduct an exploratory analysis of the primary research question that encompasses the range of core sexual risk outcomes identified in the evaluation's conceptual framework. In some cases, these "core outcomes" may not be a focus of a particular program but they may nevertheless be affected and thus merit analysis. For example, a program that encourages the use of consistent contraception might also have an impact on outcomes related to sexual activity (such as the number of partners) that are not a primary program objective. Among these core outcomes are:

- **Initiation of sexual activity**. The evaluation will focus on initiation of vaginal intercourse because of its association with teen pregnancy. However, we will also evaluate effects on sexual initiation more broadly through measures of initiation of oral and anal sex.
- **Sexual activity**. Sexual activity outcomes include frequency of sexual activity and number of sexual partners. These outcomes will be measured for vaginal intercourse, oral sex, and anal sex.
- **Unprotected sex**. Unprotected sex outcomes include use of condoms and other forms of birth control.
- **Pregnancy**. The evaluation will measure program impacts on pregnancy in all sites. However, impact studies in some sites will be underpowered to detect differences in pregnancy rates, leading us to focus on other core outcomes in the confirmatory analysis.

In some sites, measures related to one or two of the above outcomes will not be relevant, or perhaps even possible, to examine, even though they are core behaviors of interest for the overall evaluation. For example, in the two sites focused on reduction of repeat pregnancy, the evaluation will not examine measure(s) related to sexual initiation, since the target population has already been sexually active. In addition, in some sites, sensitivities among local stakeholders may constrain the set of measures that can be examined. For example, concerns by school district staff (or others) relating to asking youth about certain types of sexual activity, such as anal sex, may lead the evaluation to reduce slightly the set of measures examined.

In addition to the sexual risk outcomes listed above, the evaluation will explore estimating impacts on two other sexual risk outcomes in each site: (1) self-reported STDs and (2) births. The reason these outcomes are not listed above is simply methodological. First, self-reported STDs may be subject to considerable reporting error; we will need to conduct a careful assessment of related measure(s) before we can conclude that impacts can be rigorously estimated and thus merit being reported. Second, births are likely to be a rare outcome within the data collection period of the evaluation in all sites, including those focused on reducing repeat pregnancies. The evaluation team will have to assess the extent to which estimated impacts could emerge on this outcome before proceeding to analyze and report related impact findings.

Secondary research questions for the evaluation will be site-specific and will focus on program impacts on key mediating variables identified in the site-specific program logic models. These can include, for example, measures of knowledge about and attitudes toward sexual risk behavior and its potential consequences, including pregnancy. They could also include measures of peer influences, communication and decision-making skills, perceived behavior control, and/or access to services. The impact analysis that focuses on these questions will be entirely exploratory. That is, it will provide a detailed understanding of which, if any, of the targeted mediators a program may have affected and how these impacts may be associated with any potential effects on sexual risk behaviors. However, given the substantial "distance" in the conceptual framework between improving a given mediating variable and ultimately reducing teen pregnancy, it will provide no confirmatory evidence in any site to address the evaluation's primary research question.

2. Study Samples

In all sites, we will estimate program effectiveness using a rigorous random assignment design. If implemented correctly, random assignment ensures that the program and control groups in each site will not differ in any systematic way and, hence, that differences in mean outcomes between the two study groups will provide unbiased estimates of program impacts.

The diversity of program models and delivery settings necessitates varied random assignment approaches across sites. In selecting the most appropriate random assignment approach for each site, the following four factors were considered: (1) mode of program delivery—group- or classroom-based, individualized sessions, or a combination of the two; (2) risk of control group contamination in the delivery setting(s); (3) type of program enrollment—at defined time points or continuous; and (4) ability to generate sufficiently large samples to ensure adequate precision of impact estimation.

Table I.2 presents key features of the random assignment approach for evaluating the programs across each of the seven sites. The key features include the unit of random assignment (individual or cluster), the timing and type of random assignment (for example, rolling versus at discrete point(s) in

time), and target sample size. For four of the seven focal programs—*HealthTeacher*, *POWER Through Choices*, *Teen PEP*, and *WAIT Training* —our evaluation in the site will use a cluster randomized design, whereby clusters of youth (in schools or foster care agencies) are randomly assigned rather than individual youth. To ensure that the experimental groups in each of these clustered designs are equivalent at baseline, the clusters will be matched into small groups or pairs prior to random assignment, based on characteristics that might affect youth outcomes. For two other programs—*AIM 4 Teen Moms* and *T.O.P.P.*—the evaluation will follow a rolling random assignment design, whereby individuals are randomly assigned on an ongoing or rolling basis over many months. In these two cases, a random sequence of assignments will be generated in advance of sample enrollment to ensure an even balance of treatment and control group members throughout the enrollment period (Matts and Lachin 1988; Schulz and Grimes 2002). Finally, for the evaluation of *GEN.M*, we will also use an individual-level random assignment approach, but random assignment of youth will occur at one time for each eligible cohort (prior to each summer program cycle). More detail on the site-specific random assignment procedures can be found in Chapters II–VIII.

Table I.2. Random Assignment Approach and Target Sample Size, by Program Site

Program Model (Implementing Organization)	Unit of Assignment	Random Assignment Approach	Target Sample Size
AIM 4 Teen Moms (Children's Hospital Los Angeles)	Individuals	Rolling assignment within each program cycle using randomly permutated blocks	1,000 individuals
Gender Matters (GEN.M) (EngenderHealth)	Individuals	Simple random assignment prior to each program cycle	1,125 individuals
HealthTeacher™ Curriculum, Enhanced (Chicago Public Schools)	Schools	Matched-pair random assignment prior to 2010-2011 academic year	17 middle schools
POWER Through Choices (Oklahoma Institute for Child Advocacy)	Group foster homes	Matched pair random assignment within state for program cycle	40 group homes (up to 120 cohorts of foster care youth across the 40 homes)
Teen Options to Prevent Pregnancy (T.O.P.P.) (OhioHealth Research and Innovation Institute)	Individuals	Rolling assignment over 18 months using randomly permutated blocks	600 individuals
Teen PEP (Princeton Center for Leadership Training)	Schools	Matched-pair random assignment prior to program delivery (timing varies across schools)	16 high schools
WAIT Training (Live the Life)	Schools	Stratified random assignment prior to the spring 2012 semester	16 school clusters, including both middle and high schools

3. Data Collection and Timeline

The analysis of program impacts will rely primarily on survey data collected as part of the PPA evaluation. In most sites, data will be collected from the study sample through a baseline and two follow-up surveys. For the *POWER Through Choices* and *T.O.P.P.* evaluations, three follow-up surveys will be administered. The timing of the follow-up surveys will vary across sites in accordance with each program's theory of change and implementation schedule (Table I.3), and is discussed in more detail in Chapters II–VIII. The timing of the final follow-up survey will range from 13 to 30 months after baseline, or from 12 to 24 months after program completion. (A detailed program implementation and data collection schedule for each site can be found in Appendix A.)

Table I.3. Data Collection Waves, Timing, and Mode, by Program Site

Program Model (Implementing Organization)	Length of Program	First Follow-Up	Second Follow-Up	Third Follow-Up	Mode of Data Collection
AIM 4 Teen Moms (Children's Hospital Los Angeles)	12 weeks	12 months after baseline	24 months after baseline	---	Audio Computer-Assisted Survey Instrument (ACASI)
Gender Matters (GEN.M) (EngenderHealth)	5 days	6 months after baseline	18 months after baseline	---	Paper and pencil (baseline) Web-based and telephone (follow-ups)
HealthTeacher™ Curriculum, Enhanced (Chicago Public Schools)	16 weeks	9 months after baseline	14 months after baseline	---	Paper and pencil (baseline) Paper and pencil and telephone (follow-ups)
POWER Through Choices (Oklahoma Institute for Child Advocacy)	5–12 weeks	5-12 weeks after baseline (post-test)	7–9 months after baseline	13–15 months after baseline	Paper and pencil
Teen Options to Prevent Pregnancy (T.O.P.P.) (OhioHealth Research and Innovation Institute)	18 months	6 months after baseline	18 months after baseline	30 months after baseline	Paper and pencil (baseline) Telephone (follow-ups)
Teen PEP (Princeton Center for Leadership Training)	5–16 weeks	9-12 months after baseline	21-24 months after baseline	---	Paper and pencil Paper and pencil and telephone (follow-ups)
WAIT Training (Live the Life)	2 years, 8 days each year	20 months after baseline	32 months after baseline	---	Paper and pencil Paper and pencil and telephone (follow-ups)

The baseline and follow-up survey instruments for each site are being developed in two stages. The first stage, completed in September 2011, involved the development of "concordance" baseline and follow-up instruments whose core items are to be used in the PPA evaluation and other HHS-funded teen pregnancy prevention evaluations. In 2010 and 2011, the PPA evaluation team collaborated closely with OAH, ACYF, and other members of a federal working group to modify a previously drafted PPA baseline instrument into a concordance instrument and to develop a concordance follow-up instrument.[1] The resulting concordance instruments include three types of questions: (1) a core set of demographic variables and outcome measures that are to be included in all evaluations of federal teen pregnancy prevention grant programs; (2) a defined set of performance measures on which all federal grantees are required to report; and (3) a more general set of covariates and mediators that can be tailored to different program models, target populations, and settings. The Office of Management and Budget (OMB) approved the baseline and follow-up concordance instruments in August 2011 and September 2011, respectively.[2]

The second stage of the instrument development process involves developing the concordance instruments into site-specific instruments for each PPA site, in collaboration with local site partners. The site-specific instruments include items unique to each site, but also keep the demographic variables and outcome measures specified as core in the concordance instruments. OMB approval for the seven site-specific baseline instruments was received in August 2011. The PPA evaluation team continues to work closely with the local site teams to develop the site-specific follow-up instruments. In most sites, the baseline survey will be a self-administered paper-and-pencil questionnaire, and follow-up data will be collected again through a paper-and-pencil questionnaire and/or by telephone interviews; in one case, a web-based survey may be used (Table I.3).

D. Analytic Approach and Reporting

Our basic approach to estimating impacts will be to compare the outcomes of members of the program group with the outcomes of members of the control group. More specifically, our analytic plan has four main components: (1) an early analysis of baseline data, (2) a primary impact analysis of key sexual risk outcome measures, (3) exploratory analyses of secondary research questions, and (4) reporting of results.

1. Baseline Analysis

As soon as baseline data collection has been completed in each site, we will begin preliminary data analyses. This early analysis of baseline data will achieve three goals. First, we will use the data to describe the study sample and compare it with the target population identified in the site-specific program logic model. Second, we will assess whether random assignment successfully generated

[1] The original PPA baseline instrument was approved by the Office of Management and Budget (OMB) in July 2010. This initial baseline instrument was used in the *Health Teacher*/Chicago site, where program implementation began in fall 2010, before the concordance baseline instrument had been finalized and approved by OMB. In the other six sites, the baseline instrument used was a site-specific version of the concordance instrument, as described above.

[2] Emergency approval for the concordance and site-specific baseline instruments was received in August 2011. A request for full clearance for the baseline instruments was submitted to OMB in September 2011.

program and control groups balanced on important baseline characteristics. Third, we will examine the data for problem spots on the survey instruments (for example, patterns of incomplete or inconsistent responses) and use this information to help improve the follow-up surveys.

2. Primary Impact Analysis

To answer the primary research question about each site's program impacts on key sexual risk outcomes, we will use a well-tested approach based on multivariate regression models. With a random assignment design, unbiased impact estimates can be obtained from the difference in unadjusted mean outcomes between the treatment and control groups. However, we can improve the precision of the estimates by using regression models to control for covariates, especially baseline measures of outcomes. Regression adjustment can also account for any strata or blocking variables used in conducting random assignment, or for any differences between the program and control groups in baseline characteristics that arise by chance or from survey nonresponse.

The empirical specification for the model will depend on the unit of random assignment. With random assignment of youth, our model can be expressed as

$$(1) \quad y_i = \beta' x_i + \lambda T_i + \varepsilon_i$$

where y_i is the outcome of interest for youth i; x_i is a vector of baseline characteristics; T_i is an indicator equal to one for youth in the treatment group and zero for youth in the control group; and ε_i is a random error term. The vector of baseline characteristics x_i will include demographic characteristics such as age, gender, race/ethnicity, and baseline measures of the outcomes. The parameter estimate for λ is the estimated impact of the program.

In sites where clusters (schools or foster care homes), rather than individual youth, are the unit of assignment, the estimation must account for the correlation of outcomes among youth in the same cluster, as they will all be randomly assigned as a single unit, and each sample member cannot be considered statistically independent. To account for this dependence, we can modify the previous regression model as

$$(2) \quad y_{is} = \beta' x_{is} + \lambda T_{is} + \eta_s + \varepsilon_{is} .$$

The general structure of the model is the same, but now y_{is} is the outcome measure for individual i in cluster s (and similarly for the treatment status indicator, T_{is}, vector of baseline characteristics, x_{is}, and the error term ε_{is}). Most important, the error term in Equation (2) accounts for the clustering of youth within clusters because of the inclusion of the cluster-level error term η_s—a cluster "random effect." If this error term is excluded, the precision of the impact estimates could be seriously overstated. As in Equation (1), the estimated impact of the program is λ.

Equation (1) or (2) will be estimated separately for each primary outcome in each site. The specific method for estimating the parameters of the models will depend on the form of the dependent variable. Ordinary least squares will be used for continuous variables (such as number of partners), whereas logistic regression procedures will be specified for binary outcomes (such as ever had sexual intercourse).

In each program site, weights will be created and applied to individual sample members that account for any differences in the probability of assignment to the two experimental groups and/or

the probability of nonresponse of the follow-up survey(s). Using these weights will ensure that the study sample that we analyze in each site reflects the observed characteristics of the full study sample, regardless of assignment status or possible attrition. For the four program sites with clustered designs, we will further weight sample members so that each cluster is given equal weight (that is, has equal weighted sample size) in the analysis. This approach will lead the impact estimate to reflect the average cluster (school or foster home) in the study sample, which is the most relevant estimate for programs designed to serve youth at this cluster level. However, we will also run sensitivity analyses that do not apply this weight, effectively weighting each cluster in the analysis by its nominal sample size.

We will control for multiple hypothesis testing. The chance of falsely identifying an impact as statistically significant increases when effects on many outcomes are examined, so we will limit the primary research question for each site to a small set of key sexual behavior outcomes. Within this small set of key outcomes, we will also consider applying a formal statistical correction for multiple hypothesis testing, such as the Benjamini-Hochberg (Benjamini and Hochberg 1995) or Tukey-Kramer (Tukey 1953; Kramer 1956) method. We will discuss the pros and cons of these adjustments in a subsequent general analysis plan memo.

3. Analysis of Secondary Research Questions

As described in Section I.C, we will also define and address the following secondary research questions for each site:

- **Subgroup analyses.** To examine whether the programs were more effective for some youth than for others, we will estimate impacts for subgroups of youth by adding a term to Equations (1) and (2) that interacts the treatment indicator with a binary indicator of a particular subgroup. The regression coefficient on this term will provide an estimate of the difference in the program effect across the subgroups. Subgroups of particular interest include gender and baseline sexual experience—the two priority subgroups defined for the HHS PPRER. For sites where the program model being tested has a specific focus on subgroup impacts (for example, delaying sexual initiation among youth sexually inactive at baseline), such subgroup analyses might be elevated to the primary impact analysis.

- **Impacts on mediating variables.** Our analysis of data from the short-term follow-up survey will focus on estimating program impacts on key mediating or intermediate outcomes specified in the program logic models. We will estimate impacts on these outcomes following the same approach described in Equations (1) and (2). We will use findings for descriptive purposes and to help us interpret our primary impact analysis, not to draw overall conclusions about program effectiveness.

- **Variation in impacts by participation levels.** Our primary impact analysis will include the full study sample, yielding intent-to-treat (ITT) estimates that do not account for varying participation rates among youth assigned to the treatment group. As exploratory analyses, we will consider adjusting for participation levels in two ways. First, to account for youth who do not attend *any* program sessions or activities, we can make the standard Bloom adjustment (Bloom 1984) to calculate estimates of the treatment on the treated (TOT). In addition, to explore the association between program dosage—the degree of program participation—and impacts, we can conduct propensity score analyses, whereby youth with the different levels of program attendance are matched to a

subset of control group youth with similar demographic and baseline characteristics. This method will yield an estimate of program impacts for youth with different levels of program attendance.

4. Reporting

We will produce a final impact analysis report for each site. These independent, site-specific impact reports will address the key impact evaluation questions, synthesize findings from the impact and implementation analyses, and provide interpretations of the findings that are useful and accessible to a broad audience. Due to differences in the program implementation and data collection schedules across sites, the timing of the final impact study reports will vary by site, but will follow the completion of survey data collection in each site.

In addition to this final round of analysis and reporting, we will conduct an interim impact analysis that will draw on just the baseline and first round of follow-up data. We will combine the findings from the site-specific analyses into a single report, which we will complete in late 2014. It is possible that, for one or two sites, data collection may not be complete at the time of this interim analysis and so the findings may be based on only a partial sample. Hence, these interim analysis findings will be preliminary, and we will emphasize that in the report.

II. EVALUATION OF *AIM 4 TEEN MOMS*

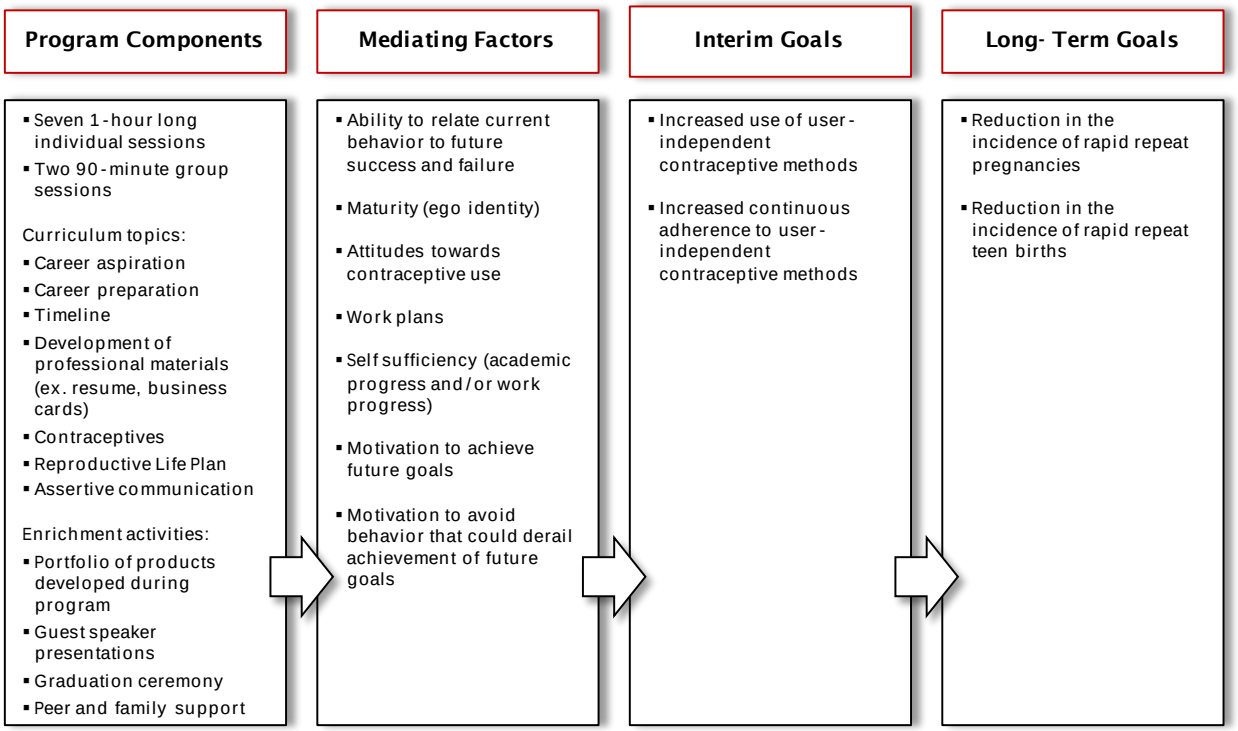
The rate of repeat pregnancies and births among teen mothers is alarmingly high. In 2009, 23 percent of births among women ages 15 to 19 were second or higher order (National Center for Health Statistics 2011). Teen mothers who bear a second child soon after the birth of their first child are a particularly vulnerable population. They are less likely to complete high school or obtain a GED and less likely to be working and become self-sufficient adults compared to teen mothers who delay their second birth until their 20s (Klerman 2004). *AIM (Adult Identity Mentoring) 4 Teen Moms* is a new, innovative but largely untested program that targets new teen mothers and aims to reduce the occurrence of high-risk behaviors that lead to repeat pregnancies. The program is an adaptation of *Project AIM*, a school-based curriculum targeting middle school students with evidence of effectiveness (Clark et al. 2005). *AIM 4 Teen Moms* focuses on protective factors, and encourages and guides teen mothers to envision and pursue a positive future for themselves and their child. Mathematica, in collaboration with ETR Associates and Children's Hospital of Los Angeles (CHLA), is conducting an experimental evaluation of *AIM 4 Teen Moms* with a focus on its effectiveness in reducing rapid repeat pregnancy among parenting teens.

A. Program Features and Evaluation Setting

As with the original *Project AIM* program, the logic model for *AIM 4 Teen Moms* (Figure II.1) is derived from the Theory of Possible Selves, Theory of Hopelessness (Boland et al. 2005), and positive youth development principles. Drawing on these theories, the program provides an opportunity to: (1) develop a clear vision of a positive future; (2) engage in skill building activities, such as goal-setting, planning, and communication; and (3) connect current behaviors with future success or failure. In adapting the program to target teen mothers, *AIM 4 Teen Moms* has been modified to include sections that are relevant to parenting teens, such as information on birth control options, reproductive life plans, parenting skills, and the use of experience in parenting as a strength to motivate youth. In addition, the curriculum has been adapted for older teens between the ages of 15 and 19, the target population of the study.

The program is provided by trained advisors and delivered in seven 60-minute individual (that is, advisor-to-participant) sessions and two 90-minute group sessions over a 12-week period. Individual sessions are generally held in teens' homes, as this setting tends to be most convenient for teen parents and eliminates common barriers to participating, such as transportation or child care. The first group session is held after the completion of the first four individual sessions, and the second is held after the completion of all individual sessions. Group sessions are planned to include a minimum of 4-15 participants. They are facilitated by one or two advisors and are held in locations easily accessible by teens, such as a conference room at CHLA.

Figure II.1. Logic Model of the AIM 4 Teen Moms Intervention



The goal of the sessions is to encourage new teen mothers to envision a positive future for themselves and their children and to develop a plan to achieve it. Advisors help participants identify a career that would allow them to become successful and self-sufficient adults. They then help participants develop and begin to implement a step-by-step plan to embark on their desired career, including preparing resumes, honing interviewing skills, and obtaining the necessary qualifications. Participants are also encouraged to think about how their reproductive choices can advance or hinder progress toward their goals. Using information about birth control options and other elements of a healthy lifestyle, they create a Reproductive Life Plan that is compatible with their aspirations. To help follow the plan, teens participate in various interactive exercises to improve their negotiation and communication skills, identify sources of support, and build self-confidence.

The objective of the group sessions is to reinforce the goals and skills developed during the individual sessions. The group sessions are designed to bolster the teens' confidence, motivation, and knowledge. They provide an opportunity to share aspirations and receive positive feedback and support from peers, and to practice negotiation and communication skills through role-playing activities. Guest speakers who have gone through similar experiences share their insights and answer questions. The final group session features a graduation ceremony, where family and friends can show their support for the progress each teen mother has made toward defining and realizing a positive future.

Evaluation Sample and Setting

The evaluation of *AIM 4 Teen Moms* will be conducted in Los Angeles County and will focus on a sample of young women, ages 15 to 19, who have had a child within the past six months and are not currently pregnant. We plan to recruit approximately 1,000 young women, half of whom will be randomly assigned to the program group and will have the opportunity to participate in *AIM 4 Teen Moms*. Enrollment in the study will take place on a rolling basis over a two-year period, and those

selected for the program group will begin to participate in *AIM 4 Teen Moms* soon after study enrollment.

Original plans called for recruiting of young women into the study sample at two sites:

- *El Nido Family Centers.* The Family Centers are located in Los Angeles County's least privileged Service Planning Areas, including South Los Angeles (the largest), Mission Hills, and Pacoima. The centers offer support to teen parents, such as linking pregnant and parenting teens with services that are intended to help them succeed in the future, such as educational assistance, vocational training, housing, and nutritional or income support.
- *CHLA's Project NATEEN (Networking, Advocacy, Teaching, Education, Empowerment & Nutrition).* Housed within the Division of Adolescent Medicine at CHLA, Project NATEEN aims to prepare teen parents for employment and promotes future planning and independence.

However, given that the number of sample-eligible young women at these sites is lower than projected, recruitment will likely be expanded to additional sites in Los Angeles. These include organizations providing WIC-based services, which are attractive both because they may serve large numbers of eligible young women and because they do not offer services (such as access to contraception) that overlap with the aims of the program; high schools for pregnant or parenting teens; one residential home and their transitional housing program for teen mothers; and health fairs and other events in the community. While this expansion may change the composition of teen mothers who participate in the evaluation, it will not change the fundamental target population for the study—young women, ages 15 to 19, who reside in Los Angeles and have had a child in the past six months.

Teen mothers recruited from these sites are likely to be at considerable risk for a repeat pregnancy. The areas served by both the El Nido Family Centers and CHLA's Project NATEEN have relatively high rates of teen births and STDs, in addition to other risk factors such as crime, unemployment, and low educational attainment (Los Angeles County Department of Public Health 2009). This is particularly true of South Los Angeles, which will likely be the largest source of eligible women for the study. For example, in 2009, 28 percent of households in South Los Angeles had income below the federal poverty level, compared to 12.5 percent in the United States and 16 percent for all of Los Angeles County. South Los Angeles also has the largest immigrant population in the county—45 percent of the population is foreign-born, and has a very large Latino population (64 percent of all residents), a racial/ethnic group with the highest teen birth rate in the country.

B. Evaluation Design

Consistent with a random assignment design, the fundamental difference between the treatment and control conditions is that treatment group members will have the opportunity to participate in *AIM 4 Teen Moms* (Table II.1). In turn, the evaluation will compare outcomes of teen mothers with access to *AIM 4 Teen Moms* and other existing services in the community against outcomes of teen mothers with access only to existing services. There is no equivalent or alternative program or curriculum offered to members of the control group; however, the evaluation takes place in a setting where other pregnancy-related or teen parent services are available to eligible women. For example, women in both the treatment and the control group have the opportunity to access services offered by the centers, as described in Table II.1. In addition, many women may be able to access services

related to pregnancy prevention that are available in Los Angeles County, outside the Family Centers.

Table II.1. Summary of Contrast Between the Treatment and Control Conditions *(AIM 4 Teen Moms)*

	Treatment Group	Control Group
Intervention	*AIM 4 Teen Moms*	No intervention
Existing Services	- Services offered by El Nido Family Centers and Project NATEEN, such as educational assistance, housing assistance, vocational training, preparation for future employment, prenatal care, nutritional and income support - Other services and programs available in LA County related to teen pregnancy prevention (will be assessed through participants surveys, focus groups, and an in-depth implementation study)	

The frequency with which women access these kinds of "existing services" will be important to measure and consider as part of the analysis of findings from the evaluation. Information about service receipt will come from two main sources: (1) questions on follow-up surveys conducted as part of the impact study, and (2) three focus groups conducted as part of an in-depth implementation study site visit to understand the program and its context—two with members of the treatment group and one with members of the control group. In addition, interviews with program stakeholders and other key informants, also conducted as part of the site visit, will help describe the backgrounds and experiences of young women served by *Aim 4 Teen Moms* and the kinds of services and activities that they may access in the community.

Sample Enrollment and Random Assignment

Aim 4 Teen Moms will be offered seven times over a two-year period. Sample enrollment will occur approximately six weeks prior to program delivery for each program cycle, and will follow a two-step process. First, potential participants will be screened for eligibility. Second, eligible young women will be assigned to field staff who will collect active consent for participation in the study and will administer the baseline survey. In most cases, consent will be gathered at the women's residence, allowing baseline data to be collected at the same time as consent. In order to be included in the evaluation sample and randomly assigned, young women must both provide active consent and complete the baseline survey. Women who meet both criteria will be contacted by program staff prior to random assignment—approximately 1-2 weeks before delivery of the first program session—to make sure that they are still interested in participating in the study and the program. This step reduces the risk of post-random assignment attrition, which can undermine the rigor of the evaluation.

Assignment to the program or control groups will be based on a random allocation sequence (known as permuted block randomization), with stratification by recruitment site.[3] Under this method, a random sequence of treatment and control assignments is generated in advance of sample enrollment within each stratum, with the sequence constrained to achieve a roughly constant allocation of treatment ("T") and control ("C") assignments throughout the enrollment period. To

[3] Stratification will be based on location and proximity of the centers and is necessary for the appropriate allocation of study participants among the advisors.

achieve this constant allocation, the assignment sequence is generated by first creating smaller "blocks" of assignments of varying lengths (for example, "TC" or "TTCC"), which are then combined to create the full randomization sequence. For this study, we will use a 1:1 allocation ratio to ensure a balanced sample, and employ random block sizes up to four characters. Separate randomization lists will be created for each recruitment site. The allocation ratio could change to ensure an appropriate sample size for the group sessions; for example, the ratio of young women assigned to the program could increase if recruitment is slower than anticipated. This change would be modest, however, leading to little, if any, meaningful reduction in statistical power.

Measuring Program Impacts

Impacts will be analyzed approximately 12 and 24 months after women complete the baseline survey, or about 9 and 21 months after women selected for the program group are scheduled to complete *Aim 4 Teen Moms*. The analysis at 12 months will focus primarily on measures of mediating factors (intermediate outcomes) and sexual risk behavior, while the analysis at 24 months will focus primarily on sexual risk behavior and repeat pregnancies. Data on participants from all enrollment sites will be pooled for primary analysis.

Given the study's random assignment (experimental) design, we will be able to calculate valid impact estimates at these two time points by simply comparing outcomes between women in the treatment and control groups. However, in order to increase precision of the impact estimates, we will use regression models to account for differences in baseline demographic characteristics, attitudes, perceptions, knowledge, behaviors, and other outcomes between women in the two experimental groups that may arise by chance.

Main impact estimates will be based on an intent-to-treat (ITT) analysis, in which all sample members are included in the analysis regardless of their participation in the program. This approach will yield an estimate of the program's average impact among young women given the opportunity to participate in *AIM 4 Teen Moms*, which is the most relevant estimate for a voluntary program. The magnitude of the ITT estimates could be affected by incomplete participation or nonparticipation by treatment group members and/or by participation of control group members in other pregnancy prevention services, making these factors important to consider in interpreting the findings. To identify these factors, advisors will track attendance in each session for members of the treatment group, and data from all sample members on the receipt of other pregnancy-related services will be collected through the follow-up surveys and focus groups.

Data

Data for the impact analysis will be obtained from surveys administered at three points in time: (1) at baseline, shortly before women in the program group begin participating in *AIM 4 Teen Moms*; (2) at first follow-up, approximately 12 months after baseline; and (3) at second follow-up, approximately 24 months after baseline. These surveys will be administered in participants' homes using audio computer-assisted self-interview (ACASI) technology, to accommodate the sample's low anticipated literacy level. To assist in the survey administration, trained Mathematica field staff will provide the survey to each sample member on a laptop and remain available in the home at the time of the data collection to provide support if needed. Incentives will be provided for survey completion, in the amounts of $20 for the baseline and 12-month survey and $30 for the 24-month survey. The incentives are especially important for parenting teens, who may find it especially challenging to make time to complete the surveys.

Outcome Measures

Drawing on the survey data, the study team will construct, and estimate impacts for, a range of outcome measures (see Table II.2). These measures fall into two broad categories: (1) sexual risk outcomes, including measures and consequences of sexual behavior, most notably repeat pregnancy, and (2) intermediate outcomes, the mediating factors through which the program would most likely impact sexual risk behavior (see Figure II.1).

Table II.2. Planned Outcomes for Measuring Program Impacts *(AIM 4 Teen Moms)*

Sexual Risk Outcomes	Intermediate Outcomes
Sexual Behavior: - *Prevalence of unprotected sex* - Number of sexual partners - Prevalence of sexual activity **Consequences:** - *Repeat pregnancy* - Births - STDs	- Intention to use contraceptives - Preferred timing of second child - Attitudes about contraceptives - School enrollment and academic progress - Academic aspiration - Employment status - Living circumstances - Possible Future Selves - Motivation to achieve future goals (Regulatory Focus Questionnaire Promotion and Prevention score)

Note: Outcomes shown in italics reflect the confirmatory outcomes for the impact analysis.

Two of the sexual risk outcomes shown in Table II.2 will serve as "confirmatory" outcomes of program impacts. The first, incidence of repeat pregnancy, will be measured as the proportion of sample members who report a pregnancy between baseline and the final (24-month) follow-up, and will serve as the primary outcome for evaluating whether the program was successful in reducing the risk of subsequent pregnancies among teen mothers. This outcome is naturally the most central to the evaluation; however, as described below in more detail, the anticipated power to detect an impact on repeat pregnancy may be marginal. For this reason, the study will also focus on evaluating the program's success in increasing the use of contraceptives, which would potentially reduce the risk of a repeat pregnancy. Thus, the second confirmatory outcome will be prevalence of unprotected intercourse. This outcome will be measured in two ways: (1) as the proportion of the sample that had intercourse without the consistent use of any (effective) contraceptives, and (2) as the proportion of the sample using user-independent methods. Although the second measure is a subset of the first one, it is vital to our understanding of the program's impact because *AIM 4 Teen Moms* promotes the use of user-independent methods and an impact on this outcome may be obscured within the broader measure of the use of any effective contraception.

Beyond this narrow, confirmatory impact analysis, a comprehensive exploratory analysis will estimate and assess impacts across a range of other outcomes shown in Table II.2—including the remaining sexual risk outcomes (shown in the left column) and the full set of intermediate outcomes (shown in the right column). Although these outcomes will not be the basis for confirmatory evidence on the effects of *AIM 4 Teen Moms*, they will provide a complete picture of potential effects of the program on teen sexual risk behaviors and their consequences, and pathways through which the program may affect sexual behaviors. Findings from the analysis of mediating factors will help us understand the sources and nature of any confirmatory evidence of program impacts on unprotected sex and pregnancy. Examples of possible mediators include shorter-term outcomes such as attitudes toward and intention to use contraceptives and motivation to achieve future goals, as well as longer-term outcomes such as employment status and academic progress that can be best examined at final follow-up.

Sample Size and Statistical Power

As a means of illustrating the study's power to detect program impacts on the confirmatory outcomes, Table II.3 displays the estimated minimum detectable impact (MDI) for two measures: (1) prevalence of user-independent contraceptives by first follow-up; and (2) the incidence of repeat pregnancy by second follow-up. For the former, the estimated MDI is an increase of around 6 or 7 percentage points from a control group mean of 20 percent, which is within the range detected by programs focusing on reducing sexual risk behavior with demonstrated evidence of effectiveness (Coyle et al. 2006; DiClemente 2009; Jemmott et al. 2005).[4] For the latter outcome, repeat pregnancy, the estimated MDI is a reduction of around 8 or 9 percentage points from a group mean of 35 percent.[5] While only nominally larger than the MDIs for the former measure, the MDIs for the incidence of repeat pregnancy reflect a more modest level of statistical power, because pregnancy is a consequence of sexual risk behavior and takes longer to emerge.

Table II.3. Minimum Detectable Impacts for Illustrative Outcomes *(AIM 4 Teen Moms)*

			Minimum Detectable Impacts	
			R^2 = High Estimate[a]	R^2 = Low Estimate[b]
	Sample Size	Mean Outcome (%)	PPΔ	PPΔ
First Follow- Up:				
Prevalence of user-independent contraceptives	850	20	6.4	7.3
Second Follow- Up:				
Incidence of repeat pregnancy	800	35	7.9	9.0

PPΔ=percentage point change. Minimum detectable impacts are estimated for a 5 percent significance level using a two-tailed test and 80 percent power. The equation used to calculate the minimum detectable impacts for individual-level random assignment is:

$$MDI = 2.8 \times \sigma \sqrt{\rho \times \left(1 - R_i^2\right) \times \left(\frac{1}{N_T} + \frac{1}{N_C}\right)},$$

where σ is the standard deviation of the outcome measure, and NT and NC are the number of individuals in the treatment and control groups. The estimated sample size is based on 85 percent and 80 percent retention rates for the first and second follow-ups, respectively.

[a]High estimate of R-squared is 0.3.

[b]Low estimate of R-squared is 0.1.

[4] Prevalence of user-independent contraceptives is defined as the likelihood of using such methods in the past 12 months. An approximate mean for the control group is estimated based on data from the 2006–2010 National Survey of Family Growth, and represents the proportion that used a user-independent contraceptive method in the past 12 months among Hispanic women ages 16-20 who reported having been pregnant at least once.

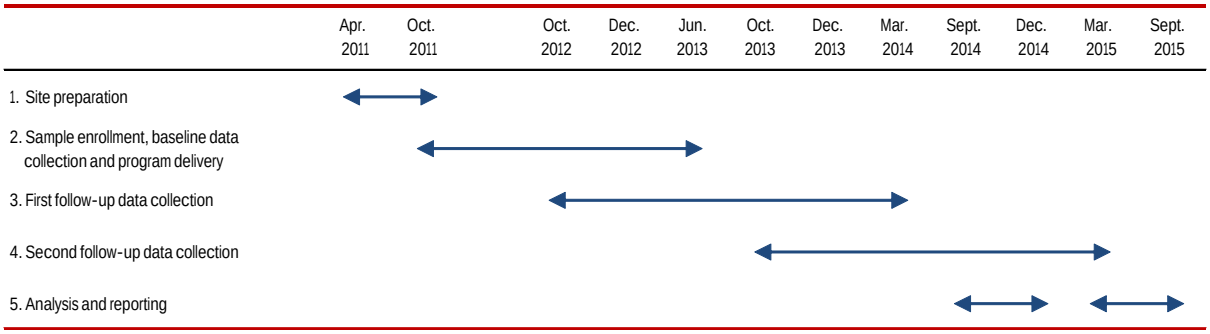
[5] Rapid repeat pregnancy is defined as the likelihood of getting pregnant in the past 24 months. An approximate mean for the control group was estimated based on birth files for 2009 published by the Centers for Disease Control and Prevention (CDC). The estimate represents the proportion of births that were second or higher order, among Hispanic women ages 17-21 residing in Los Angeles County, California.

Evaluation Timeline and Progress to Date

The study will be completed over a five-year period, beginning in the spring of 2011 (Figure II.2). During the first year of the study, the evaluation team began working with the El Nido Family Centers and CHLA's Project NATEEN to recruit a sample of teen mothers who match the target population for the program and study. In addition, the program staff hired advisors to deliver the program and the study team trained field staff to administer the survey. Sample enrollment, baseline data collection and program delivery began in fall 2011 and will continue through summer 2013. The first and second follow-up surveys will be administered 12 and 24 months after baseline for each participant, respectively. The first follow-up data collection will be completed for the full sample in summer 2014.

A final comprehensive report detailing the study findings is expected in fall 2015. The report will describe the evidence of the program's impact on the two confirmatory outcomes for the study, and will provide a comprehensive assessment of impacts on other outcomes, including both the intermediate outcomes and additional sexual risk outcomes described above. An interim report will also be released in late 2014. This report will include findings for multiple programs that are part of the overall PPA evaluation, as a means of providing OAH, and HHS more generally, with an understanding of the near-term impacts that are emerging across the programs. Findings included in this report for *AIM 4 Teen Moms* will be preliminary; they will not provide the basis for drawing any confirmatory conclusions on the effectiveness of the program.

Figure II.2. Timeline for the Evaluation of AIM 4 Teen Moms



III. EVALUATION OF *GENDER MATTERS*

Texas has the third highest teen birth rate and fourth highest teen pregnancy rate in the United States (Martin et al. 2011). In response to these troubling rankings, EngenderHealth is introducing *Gender Matters* (*GEN.M*) in Travis County, Texas, where the pregnancy rate exceeds that of the state. *GEN.M* focuses on gender norms as a determinant of adolescent sexual risk behaviors and pregnancy. The program aims to reduce teen pregnancy by challenging social constructions of femininity and masculinity associated with sexual behaviors and promoting healthy, equitable relationships. As part of the PPA evaluation, Mathematica, in collaboration with EngenderHealth and Columbia University's Mailman School of Public Health, will conduct an experimental evaluation of *GEN.M*'s effectiveness in reducing teen pregnancy and associated behavioral risks.

A. Program Features and Evaluation Setting

GEN.M is a community-based intervention with three components: (1) a 20-hour educational curriculum, (2) a social media (SMS texting and Facebook) campaign, and (3) a reunion event for *GEN.M* alumni. The social media and community event components are designed to support and reinforce key messages delivered during the curriculum-based workshops. The *GEN.M* program was developed by EngenderHealth, an international nonprofit organization focused on reproductive health issues. *GEN.M* is an adaptation of EngenderHealth's *Gender Matters* curriculum, which has been previously modified for use with young Latino and African American men in Los Angeles and 9th grade students in New York City. For *GEN.M*, the *Gender Matters* curriculum was adapted to focus specifically on the relationship between constructions of gender and associated risks for teen pregnancy.

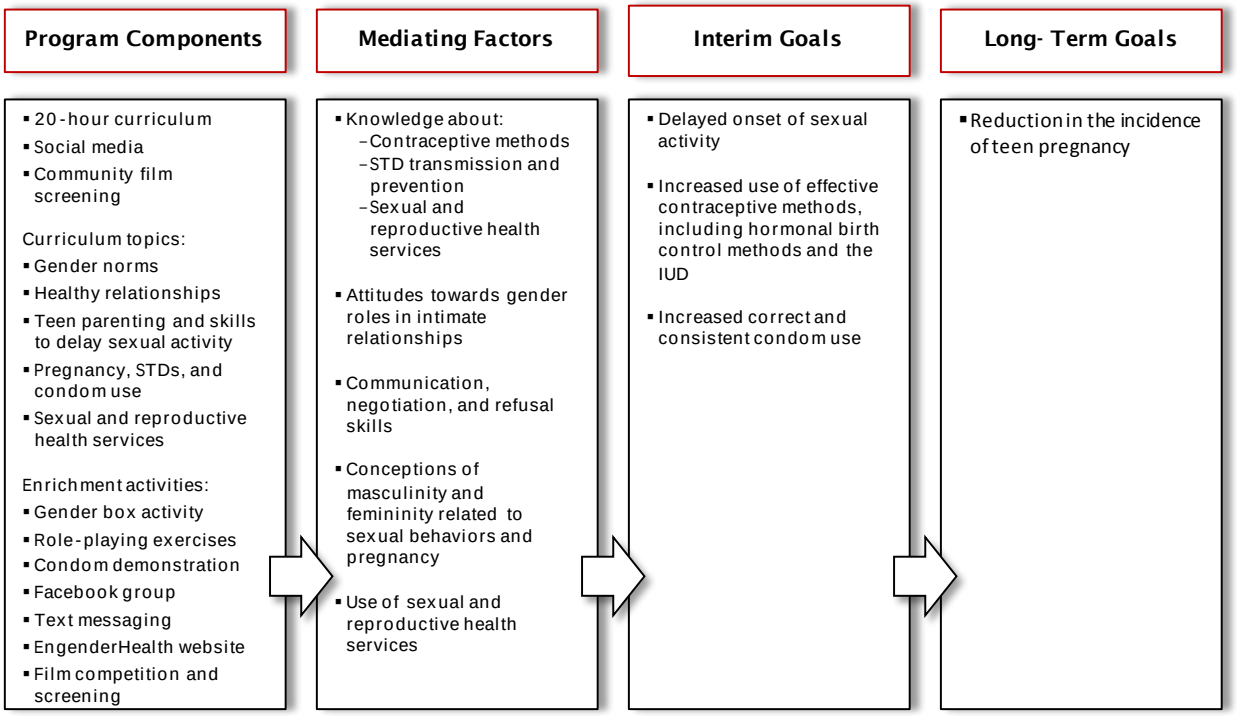
The *GEN.M* curriculum and theory of change (Figure III.1) are grounded in three psychosocial behavioral change theories. The design of the program's educational activities draws on elements of social cognitive theory, which posits that youth learn behaviors by observing peers and practicing their knowledge and skills in their own environments. The curriculum also applies the theory of gender and power to guide youth in examining how gender norms and power dynamics in relationships influence sexual risk behaviors and teenage pregnancy. Lastly, the curriculum applies social norm theory by exploring, questioning, and attempting to change perceived social norms about gender and pregnancy. The link between gender norms and sexual behavior outcomes that underlies the *GEN.M* curriculum is supported by an extensive literature. Studies have shown that adolescent males who hold stereotypical attitudes towards masculinity use condoms less consistently, have lower use of health services, and have greater belief that pregnancy validates masculinity (Kandrack et al. 1991; Pleck et al. 1993; and Courtenay 2000). Adolescent females who hold conventional views on femininity are more likely to accommodate the interests and desires of men, use condoms less consistently, and to have a lower age at first birth (Connell 1987; Stewart 2003; Ickovics and Rodin1992)

The *GEN.M* curriculum is delivered to small groups of 12 to 15 youth in 5 four-hour sessions, which will be held on consecutive days. The sessions are led by one female and one male facilitator. Although the curriculum focuses on the role of gender norms, it also addresses other determinants of teen sexual behaviors, including decision-making, communication, and negotiation skills and knowledge about effective birth control methods. The following topic areas are covered over the course of the five-day curriculum: (1) gender socialization and its influence on sexual behavior; (2) healthy relationships; (3) challenges of being a teen parent and skills to delay sexual activity; (4)

pregnancy, STDs, and how to prevent both through condom use; and (5) taking action—family planning services and individual behavior change plans to prevent pregnancy.

At the conclusion of the curriculum, participants are invited via text message to share their thoughts about the *GEN.M* program with other program participants via a private, supervised Facebook group, and to receive informational text messages. They are also invited to attend a premiere of videos that convey *GEN.M*'s messages in creative ways, created as part of a video competition among program participants. As noted above, these components of the *GEN.M* program are designed to enhance the effectiveness of the curriculum-based workshops by reinforcing messages related to progressive gender norms, healthy relationships, and pregnancy prevention.

Figure III.1. Logic Model of the *GEN.M* Intervention



Evaluation Sample and Setting

The *GEN.M* program will be tested in Austin, Texas, the capital of the state and the seat of Travis County. Youth who apply to participate in the Travis County Summer Youth Employment Program (SYEP) and who meet the study's inclusion criteria will be eligible to participate in the *GEN.M* evaluation. SYEP, a joint venture of the City of Austin and Travis County Health and Human Services departments, is a youth development program that places teens at local job sites across Travis Country for five weeks during the summer. SYEP provides teens with an opportunity to develop skills and confidence to succeed in the workplace, as well as exposure to various career options and positive adult role models. The program serves youth between the ages of 14 and 18 who reside and attend school in the City of Austin or Travis County. While the SYEP is offered to all youth attending public schools in Austin/Travis County, the program focuses recruitment on youth who are low income, at risk, or have special needs. Because SYEP caters to the same high-risk adolescent population that the *GEN.M* intervention targets, it was selected as the sample recruitment vehicle for the evaluation. The SYEP is offered twice each summer, with five-week

sessions starting in early June and mid-July. However, due to limitations in the number and size of SYEP job sites, not all SYEP applicants receive job assignments and are able to participate in the program.

Travis County as a whole faces high teen birth and STI rates, but the 12 zip codes from which SYEP primarily draws its participants are particularly high risk. These zip codes account for 80 percent of teen births and the majority of STD cases in Travis County, and have the lowest median family income in the county. Over 90 percent of this population is African American or Latino, and 55 percent of students in the county qualify for the free lunch program.

To be eligible for the *GEN.M* evaluation, interested SYEP applicants must be between the ages of 14 and 16, have obtained parental consent to participate in the evaluation, and have completed a baseline survey administered during the third week of each SYEP five-week session. Not all SYEP applicants will have received a job assignment (and, therefore, will not be participating in the SYEP) at the time of baseline administration; however, all consented, age-eligible youth will be required to complete the baseline survey in order to be enrolled in the study. The *GEN.M* curriculum will be delivered during the week after each SYEP session ends, to sample members selected to receive the *GEN.M* intervention.

B. Evaluation Design

The impact study relies on an experimental design. Members of the study sample will be randomly assigned to either a program group that receives the *GEN.M* intervention after the SYEP session, or to a control group. There is no alternative program being offered to the control group as part of this project. Thus, the evaluation will compare the outcomes of teens with access to GEN.M (and any existing services in the community) to the outcomes of teens with access only to existing services. Since the study sample will include teens enrolled in SYEP, a sizable proportion of the treatment and control group will receive the services and benefits of the SYEP program. In addition, sample members in both study groups can access existing services offered to teens in Travis County, some of which address risk factors related to teen pregnancy (Table III.1).

Table III.1. Summary of Contrast Between the Treatment and Control Conditions *(GEN.M)*

	Treatment Group	Control Group
Intervention	*GEN.M*	No intervention
Existing Services	- *Reducing the Risk (RTR)* – A 16-session evidence-based (TPP Tier 1) curriculum aimed at building skills and knowledge to prevent pregnancy, HIV, and STDs. RTR is being offered by Planned Parenthood to referred high-risk youth in some Austin high schools, with funding from a federal Office of Adolescent Health grant. - *SafePlace* – Community organization that provides adolescent intimate partner violence workshops at middle and high schools in Austin and conducts a variety of related community events.	

Some Austin high schools offer the Reducing the Risk (RTR) program to 9th and 10th grade students, some of whom can be expected to be in the *GEN.M* evaluation. This likely makes RTR the most significant source of "existing services" for youth in the evaluation, in both the treatment and control groups. Based on the past distribution of those enrolled in SYEP, roughly 15 percent of our sample is expected to participate in RTR, either before or after their enrollment in the evaluation and participation in SYEP. Follow-up surveys of *GEN.M* sample members will provide a formal measure of the extent to which the *GEN.M* sample—program and control groups—have received

pregnancy prevention program interventions in school, including RTR and any other programs or services.

As a means of further documenting the access sample youth may have to other types of pregnancy prevention-related services both in school and elsewhere, and how this contrasts with the services received in *GEN.M*, we will conduct a complementary program implementation study. This study will both provide a detailed assessment of the implementation of *GEN.M* and gather information about participation in other services and activities by members of the treatment group. Data for the study will be collected from three main sources: (1) a participant satisfaction survey administered to all program participants at the end of the *GEN.M* program, (2) in-depth interviews with select *GEN.M* participants, and (3) interviews with EngenderHealth staff.

Sample Enrollment and Random Assignment

The 20-hour *GEN.M* curriculum will be delivered to two cohorts of treatment group youth each year—after the completion of the two respective SYEP sessions—and continue for three years: 2012, 2013, and 2014. Sample intake will take place annually, between February and July of each year, for the two cohorts. Specifically, from February to April, SYEP will conduct 15-hour job readiness trainings for youth interested in SYEP. Parents of interested youth are required to attend the first session of this training to learn more about the program, to complete the program application, and to sign a consent form allowing their child to participate in SYEP. During this first session, a presentation about the *GEN.M* program and the evaluation will be made by representatives of EngenderHealth and Mathematica.[6] Parents who want their child to have a chance at participating in *GEN.M*, and agree to the child's participation in the PPA evaluation, will be asked to sign an evaluation consent form during the job readiness training.

The evaluation sample will consist of eligible SYEP applicants aged 14 to 16 years old with parental consent who complete the PPA baseline survey. Following completion of the baseline survey, youth will be stratified by age and then randomly assigned to a treatment group that receives the *GEN.M* program or the control group. Mathematica will conduct random assignment and communicate the results to EngenderHealth and the youth participating in the evaluation during the fourth week of each SYEP session..

In total, an estimated 1,140 youth will be enrolled in the study sample over the three-year period. This rough estimate is based on several assumptions. The first assumption is the size of the SYEP applicant pool and its anticipated age distribution, which should yield about 500 youth who are eligible for the evaluation each year. The second assumption is the parental consent rate among eligible applicants (assumed to be 80 percent), and the third assumption is the baseline completion rate (assumed to be 95 percent). Combining these three figures, we assume that 80 percent of age-eligible youth (or 400 youth) will obtain parental consent to participate in the evaluation each summer, and that 95 percent (380) of those youth will complete the baseline questionnaire, for a total sample across the three years of 1,140 youth.

[6] Mathematica staff will be present at the three largest job readiness trainings, which will be attended by 100 or more youth. In addition to these larger trainings, a series of smaller trainings will be held throughout Travis County between February and April. EngenderHealth staff will be present at all of the trainings.

Measuring Program Impacts

Impacts of the *GEN.M* program will be analyzed based on data collected approximately 6 and 18 months after the end of the educational curriculum, or about 7 and 19 months after youth complete the baseline survey. The 6-month analysis will focus primarily on mediating (or intermediate) outcomes, as described in Figure III.1, while the 18-month analysis will focus on sexual risk behaviors and associated consequences.

Given the random assignment design, unbiased impact estimates can be obtained by simply comparing post-intervention outcome measures between the treatment and comparison groups. To obtain more precise estimates, however, we will use regression models to control for random differences in the baseline characteristics of program and control group members and estimate regression-adjusted means of the outcomes. Whenever possible, models will include the baseline measure of a specific outcome, in addition to other risk factors. Final impact estimates will be based on pooled data from all sample members, enrolled over the three years of sample intake.

The main impact estimates will be based on an intent-to-treat (ITT) analysis that includes all youth who were randomized, regardless of their participation in the program. This approach will yield an estimate of the program's average impact among young women given the opportunity to participate in *GEN.M*, which is the most relevant estimate for a voluntary program. The magnitude of the ITT estimates could be affected by incomplete or nonparticipation by treatment group members and/or by participation of control group members in other pregnancy prevention services, making these factors important to consider in interpreting the findings. To identify these factors, program staff will track attendance during each program cycle for members of the treatment group, and data from all sample members on the receipt of other pregnancy-related services will be collected through the follow-up surveys.

Data

The data for the impact analysis will be collected through three surveys. The first is a baseline paper-and-pencil questionnaire, which will be administered on the Saturday of the third week of each SYEP session. Each of the baseline survey administrations will be overseen by a trained Mathematica interviewer, who will monitor the data collection and be responsible for distributing the questionnaires and gathering them upon completion. The second and third surveys will be administered through telephone interviews conducted by trained Mathematica interviewers at roughly 6 and 18 months after completion of the educational curriculum. To increase response rates for the two follow-up surveys, sample members will receive $20 for completing the 6-month interview and $20 for completing the 18-month interview.

Outcome Measures

Drawing on the survey data, the study team will construct, and estimate impacts for, a range of outcome measures (Table III.2). These measures fall into two broad types: (1) sexual risk outcomes, which include both measures of sexual behavior and consequences of this behavior, most notably pregnancy; and (2) intermediate outcomes, which correspond to the mediating factors through which the program would most likely impact behavior (see Figure III.1 above).

Two of the sexual risk outcomes shown in Table III.2 will serve as primary, or "confirmatory", measures of program impacts: (1) prevalence of sexual intercourse and (2) prevalence of unprotected

sexual intercourse.[7] Delaying sexual initiation and increasing the use of birth control methods among sexually active teens are key goals of *GEN.M*. Differences in the prevalence of sexual intercourse and unprotected sexual intercourse between members of the treatment and control groups will indicate whether the program was successful in achieving these goals. These primary outcome measures will be evaluated at both follow-ups, though impacts on sexual initiation are more likely to emerge at the 18-month follow-up. While the ultimate goal of *GEN.M* is to reduce teen pregnancy, any impacts of *GEN.M* on this outcome are unlikely to materialize within the 18-month follow-up period and so it cannot serve as a confirmatory outcome for the study.

Beyond this narrow, confirmatory impact analysis, a comprehensive exploratory analysis will estimate and assess impacts across the range of other outcomes shown in Table III.2, including both the remaining sexual risk outcomes (shown in the left column) and the full set of intermediate outcomes (shown in the right column). Although these outcomes will not be the basis for confirmatory evidence on the effects of the program, they will provide a complete picture of potential effects of the program on teen sexual risk behaviors and their consequences, and pathways through which the program may affect sexual behaviors. In particular, findings from the analysis of intermediate outcomes, or mediating factors, will help us to understand the sources and nature of any confirmatory evidence of program impacts on the prevalence on sexual intercourse and unprotected sex. Exploratory analyses will also be conducted to understand whether and how program impacts vary by key subgroups, such as by gender and/or race/ethnicity.

Table III.2. Planned Outcomes for Measuring Program Impacts *(GEN.M)*

Sexual Risk Outcomes	Intermediate Outcomes
Sexual Behavior: - *Prevalence of sexual intercourse* - *Prevalence of unprotected sex* - Frequency of sexual activity - Number of sexual partners **Consequences:** - Pregnancy - STDs	- Views on sexual intercourse and its importance to boys and girls - Attitudes toward communication and decision making in intimate relationships - Self-efficacy for refusing, delaying, and negotiating sex - Attitudes toward pregnancy, contraceptive use, and condoms - Knowledge about contraceptives, the risk of STDs, and risk of pregnancy - Self-efficacy about contraceptive and condom use and where to obtain sexual and reproductive health services - Use of sexual and reproductive health care services

Note: Outcomes shown in italics reflect the confirmatory outcomes for the impact analysis.

[7] In addition to the relative importance of these two outcomes, the confirmatory analysis is limited to them because of the statistical power loss associated with multiple hypotheses testing (the increased chance of falsely identifying an impact as statistically significant as the number of outcomes increases).

Sample Size and Statistical Power

As an indication of our ability to detect significant differences in outcomes related to sexual risk behavior, Table III.3 shows the minimum detectable impact (MDI) for the two confirmatory outcomes at the 18-month follow-up: (1) prevalence of sexual intercourse and (2) prevalence of unprotected sex. For the former outcome, the estimated MDI is a reduction of around 7.7 or 8.7 percentage points from a control group mean of 58 percent.[8] For the latter outcome, prevalence of unprotected sex, the estimated MDI is a reduction of around 6 percentage points from a control group mean of 15 percent.[9] As GEN.M is a unique program focusing on gender-based attitudes and behaviors as a means to prevent teen pregnancy, the existing literature provides little guidance for the magnitude of expected program impacts. Nevertheless, the estimated MDIs are within the range of program impacts found in studies of a number of other comprehensive sex education programs, suggesting they are appropriate for this evaluation (Coyle et al. 2006; DiClemente 2009; Jemmott et al. 2005).

Table III.3. Minimum Detectable Impacts for Illustrative Outcomes *(GEN.M)*

			Minimum Detectable Impacts	
			R^2 = High Estimate[a]	R^2 = Low Estimate[b]
	Sample Size	Mean Outcome (%)	PPΔ	PPΔ
Second Follow- Up:				
Prevalence of sexual intercourse	912	58	7.7	8.7
Prevalence of unprotected sex	912	15	5.5	6.3

PPΔ=percentage point change. Minimum detectable impacts are estimated for a 5 percent significance level using a two-tailed test and 80 percent power. The equation used to calculate the minimum detectable impacts for individual-level random assignment is:

$$MDI = 2.8 \times \sigma \sqrt{\left(1 - R_i^2\right) \times \left(\frac{1}{N_T} + \frac{1}{N_C}\right)},$$

where σ is the standard deviation of the outcome measure, and NT and NC are the numbers of individuals in the treatment and control groups. The estimated sample size is based on an 80 percent retention rate for the second follow-up.

[a]High estimate of R-squared is 0.3.

[b]Low estimate of R-squared is 0.1.

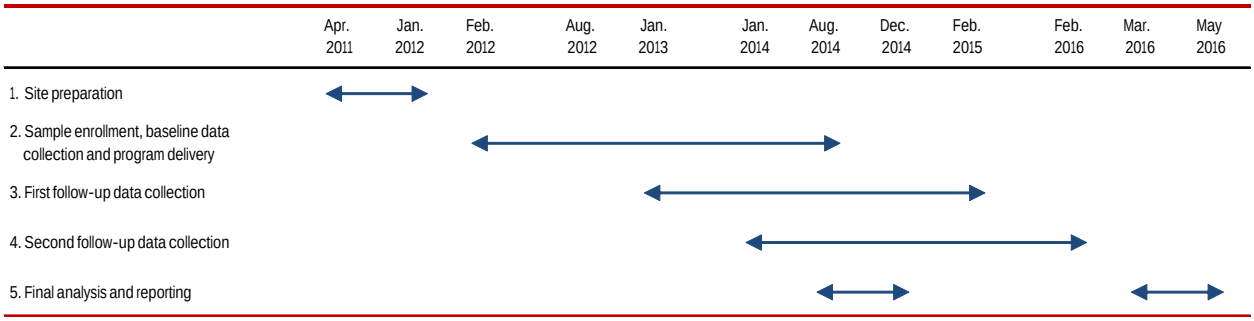
[8] Prevalence of sexual intercourse is defined as the proportion of youth who ever had sexual intercourse. An approximate mean for the control group is estimated based on data from the 2009 Youth Risk Behavior Survey (YRBS), and represents the proportion that reported ever having sexual intercourse among Hispanic and African American teens in grades 9-12.

[9] Prevalence of unprotected sex is defined as the proportion of youth who had unprotected sex during the months prior to the follow-up. An approximate mean for the control group is based on data from the 2006-2010 National Survey of Family Growth, and represents the proportion who did not use an effective contraceptive during last intercourse (in the past 12 months) among Hispanic and African American teens ages 15-19. (For this measure, teens who did not report having sex during the past 12 months were coded zero.)

Evaluation Timeline and Progress to Date

The timeline for the *GEN.M* evaluation extends from start-up planning in 2011 through final results in 2016 (Figure III.2). During the first year of the evaluation (April 2011–January 2012), the evaluation team focused on study design and logistics, including how the evaluation will be integrated into the SYEP program sequence, the development of and process for administering the baseline and follow-up survey instruments, and sample intake and random assignment procedures. Sample enrollment, baseline data collection and program delivery began in February 2012 with the start of the first of the three annual rounds of enrollment. The first follow-up data collection will occur in January and February of 2013, 2014, and 2015. Data collection for the second follow-up is expected to be completed in February 2016. A final report presenting the results of the *GEN.M* evaluation is expected at the end of the second quarter of 2016. Prior to this report, an interim report is planned for late 2014 focusing on impact findings from the first (six-month) follow-up survey. This report, which will present findings separately for each of the different PPA sites, will draw on data from the first two years of sample for the *GEN.M* evaluation.

Figure III.2. Timeline for the Evaluation of GEN.M



IV. EVALUATION OF ENHANCED *HEALTHTEACHER*

Comprehensive "off-the-shelf" health curricula offer school districts a practical option for complying with mandated requirements for teaching health education. This is particularly true of large districts, where widespread adoption of more specialized curricula may be impractical because of significant training and/or monitoring demands. As a result, there is great interest in learning whether such off-the-shelf curricula can be effective at reducing teen risk behavior. *HealthTeacher*, one of the most widely used of these curricula, is a relatively inexpensive, online health education curriculum that provides teachers with easy access to lesson plans, session materials, and links to further information. As of fall 2011, Relegent, the original developer of the curriculum, had site agreements for use of *HealthTeacher* in more than 500 school districts, including large districts in Chicago, New York City, Palm Beach and Duvall County in Florida, and Clark County in Nevada.[10]

To expand the knowledge base on the effects of affordable comprehensive sex education curricula, Mathematica is conducting an experimental evaluation of an enhanced version of the family health and sexuality module of *HealthTeacher*. The evaluation, which will focus on middle school students in Chicago Public Schools (CPS), will test whether the *HealthTeacher* curriculum is effective at delaying sexual activity and/or reducing risky sexual behavior among youth who are sexually active.

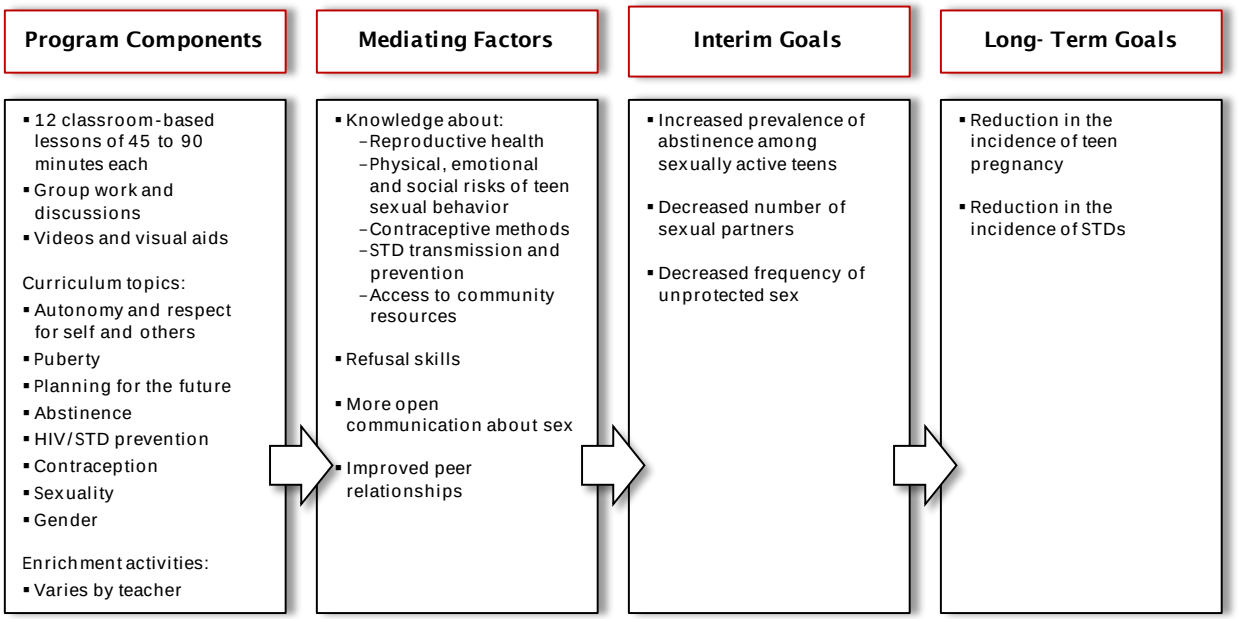
A. Program Features and Evaluation Setting

HealthTeacher is an online curriculum designed to increase students' access to medically accurate and age-appropriate information on a broad range of health topics and to equip them with the knowledge and skills necessary to avoid risk-taking behaviors (Figure III.1). The curriculum is intended for students in grades K-12, and includes different lesson plans and topics for each grade level. The lesson plan designed for middle school students consists of 10 modules, and covers topics related to substance use, nutrition, injury prevention, sexuality, mental health, and anatomy, among others. The current evaluation is designed to study the impact of an enhanced version of the "Family Health and Sexuality" module of *HealthTeacher*. Unlike the nationally available version of the module, which includes only 9 lessons, the enhanced version includes 12 lessons and meets the CPS guidelines for the content of sex education offered in 7th grade, as well as the sex education standards of the Illinois Caucus for Adolescent Health. From here on, we refer to the 12-lesson version of the "Family Health and Sexuality" module, the intervention being tested, simply as *HealthTeacher*.

The 12-lesson curriculum is considered a comprehensive sex education program and covers a variety of topics, including messages related to abstinence, contraception, making good decisions, STDs, substance use, and gender and sexuality. The lessons vary in duration from 45 to 90 minutes each, or 60 minutes per lesson on average, and are delivered by classroom teachers who received training in *HealthTeacher*. Students are active participants in the lessons. They receive handouts, complete worksheets, and engage in group activities and role-playing as a way of acquiring and demonstrating the knowledge and skills the curriculum aims to develop. In addition to the supplemental materials that are part of the lesson plans, the *HealthTeacher* website provides teachers with access to further materials and links that can be incorporated into the lessons.

[10] Source: www.healthteacher.com, accessed on March 15, 2012.

Figure IV.1. Logic Model of the *HealthTeacher* Intervention



Students are encouraged to envision specific goals for the future and consider the ways in which having unprotected sex might impact their ability to achieve those goals. Students will also discuss the emotional repercussions of teenage sex and learn how to make decisions based on their own boundaries. Lessons about puberty and the mechanics of reproduction lead into discussions of contraceptive use and STDs and pregnancy prevention, and are followed by lessons in which participants evaluate their own tolerance of differences in sexuality and gender expression and learn the importance of acceptance and respect. The lessons covering contraceptive use, sexuality, and gender were developed specifically for the CPS population. See the logic model for *HealthTeacher* (Figure IV.1) for a summary of these program components and the targeted mediating factors and intermediate and primary outcomes.

These 12 lessons and their objectives are as follows:

1. ***Recognizing Respect.*** Help students identify respectful behavior and develop communication skills that convey respect.
2. ***Changing Minds.*** Help students understand the emotional and social changes that occur during adolescence and analyze external influences, such as friends, and internal influences, such as personal growth.
3. ***Changing Bodies.*** Teach students about the female and male anatomy, including the physical changes that occur during adolescence, and help them identify reliable sources of information for questions about adolescence.
4. ***Menstruation and Sperm Production***. Teach students about the menstrual cycle and sperm production, and supply them with accurate information about female and male reproductive functions.
5. ***Looking to My Future.*** Convey the importance of goal setting for the future as a way of reducing health risks, such as STDs and teen pregnancy; help students develop skills for setting and achieving goals.

6. ***Looking at Barriers.*** Help students set goals for their future and identify barriers for achieving future goals, including sexual activity.
7. ***Abstinence.*** Convey the benefits of abstinence by helping students identify the physical, emotional, and social reasons why abstinence is a good choice for adolescents, and encourage them to advocate abstinence among their peers.
8. ***It's OK to Say No.*** Help students identify the risks associated with teen sexual behavior, and develop communication and refusal skills to avoid sexual behavior.
9. ***Preventing STD/HIV.*** Educate students about the symptoms, consequences, treatment, and transmission of HIV/STD, and help them identify community resources for testing, treatment, and information about HIV/STD in their community.
10. ***Contraceptives.*** Discuss different methods of contraception, including abstinence, condoms, hormonal methods, and user-independent methods such IUDs; help students identify all available methods, explain how to use them, and provide students with web resources for information about contraceptives.
11. and 12. ***Sexuality and Gender.*** Discuss messages and societal expectations related to Lesbian, Gay, Bisexual, and Transgender (LGBT) population; help students identify and use appropriate and accurate terminology with regard to sexual orientation and gender identity, and promote tolerance.[11]

Evaluation Sample and Setting

The evaluation is being conducted in 17 middle schools within the Chicago Public School system that serve communities with high rates of teen pregnancy or STDs. Seventh grade students in participating schools were eligible to enroll in the study; those providing active parental consent comprise the evaluation sample. The *HealthTeacher* intervention was delivered in spring 2011 to roughly half this study sample—that is, to students in 9 of the 17 schools randomly assigned to a "treatment group" (discussed below). Implementation occurred during school hours—in some schools in place of academic instruction and in others in place of regular health or physical education classes.

The 17 study schools draw students from mostly lower-income communities in three areas of Chicago: the West side, which is predominantly Hispanic; the South side, which is predominantly African American; and the North side, which is a mix of African American, Caucasian, and Middle Eastern residents. Twelve of the 17 participating schools were recruited from the West side of Chicago; the majority of the students in the sample (about 70 percent) are therefore of Hispanic ethnicity. In the 2009–2010 academic year, between 76 and 100 percent of the students who were enrolled in participating schools qualified for free or reduced-price lunch programs.[12]

[11] The lessons use sections from the "Dealing with Difference: Opening Dialogue About Lesbian, Gay & Straight Issues" curriculum from Human Relations Media.

[12] Source: Institute for Education Science's National Center for Education Statistics.

B. Evaluation Design

Following sample recruitment, the 17 study schools were randomly assigned to either a treatment group that implemented the *HealthTeacher* curriculum in 7th grade, or a control group that did not implement *HealthTeacher*. As shown in Table IV.1, the main difference between the treatment and control conditions is that youth in the treatment group will receive the *HealthTeacher* curriculum. There was no equivalent or alternative program offered to 7th graders in the control schools during the 2010–2011 academic year as part of the study. Thus, the evaluation will compare the outcomes of students who received *HealthTeacher* (and had access to any other existing services in the community) to the outcomes of students with only access to existing services. Based on data collected as part of the PPA implementation study in this site—which included teacher and staff interviews and focus groups with students—currently, there appear to be few school- or community-based services providing sexual education or related services to middle school-aged youth. As a result, the study schools serve as an excellent setting for conducting a meaningful impact study of the *HealthTeacher* curriculum.

Table IV.1. Summary of Contrast Between the Treatment and Control Conditions *(HealthTeacher)*

	Treatment Group	Control Group
Intervention	*HealthTeacher*	No intervention
Existing Services	- Schools in the study did not implement any other sex education programs in 7th grade in academic year 2011–2012. Schools are committed to not introducing any new curriculum before the end of the evaluation (after completion of second follow-up data collection). - Based on findings from teacher interviews and focus groups, there are no known after-school or community-based programs related to sex education.	

Sample Enrollment and Random Assignment

Enrollment of the study sample involved two steps: (1) recruitment of schools in high-risk Chicago communities, and (2) enrollment of eligible students in participating schools. Using data from the Chicago Department of Public Health, the evaluation team first selected Chicago Community Areas with high rates of teenage births or STDs. Within the selected communities, the team then identified more than 60 middle schools that expected to enroll at least 75 students in 7th grade in the 2010–2011 academic year. Subsequent outreach to principals and staff at these schools identified a total of 19 schools with interest and ability to participate in the study. Of these 19 schools, 2 later dropped out of the study because consent could not be gathered in time for program implementation, yielding a final evaluation sample of 17 schools.

The eligible study population in each school included students in the 7th grade who were enrolled in the course in which *HealthTeacher* would be provided (if the school was assigned to the treatment group). In most schools, this course was mandatory for students so the eligible sample included all students in 7th grade. Enrollment of students in the evaluation took place between September and early November of 2010. The evaluation team sought active consent from parents or legal guardians specifically for participation in the evaluation; passive consent for participation in a health curriculum was covered under a blanket consent distributed by schools at the beginning of the school year. Mathematica offered small classroom incentives (for example, a class pizza party) to encourage the return of forms.

Random assignment occurred at the school level. To increase the chance that the sample is balanced on important baseline characteristics, participating schools were matched into pairs, and randomly assigned to either the treatment or control group within each pair. Matching was based on school size, race/ethnicity of the student population, and the proportion of students receiving free or reduced-price lunch. Because an odd number of schools was recruited, one of the schools was randomized without being matched. Assignment occurred in late November 2010, after completion of consent gathering in 17 schools.

The resulting study sample included a total of 1,460 students across the 17 schools—9 schools in the treatment group and 8 schools in the control group. The student sample reflected a 70 percent consent rate, meaning 70 percent of those eligible to participate in the study across the 17 schools gained active parental consent to do so.

Measuring Program Impacts

Impacts of *HealthTeacher* will be analyzed approximately 9 and 14 months after the program start—or 6 and 12 months after the end of program delivery. All estimates will be based on an intent-to-treat (ITT) analysis, meaning the analysis sample will consist of all youth in the study sample, including those in the treatment group who may ultimately not participate in the program. This approach will provide an estimate of the average program impact among all program-eligible students (who are participating in the study), which is the most relevant estimate for a mandatory in-school program.

Given the random assignment design, valid impacts can be estimated by simply comparing the post-intervention outcomes between the treatment and control groups. However, in order to increase precision of the impact estimates, regression models will be used to measure program impacts, using baseline data on students' characteristics, reported behaviors, and other information to form covariates. To account for the possible correlation in the outcomes among students within the same school (intraclass correlation), we will use the Huber-White method to estimate the standard error associated with each impact estimate.

Data

The analysis of program impacts will be based on three rounds of self-reported data, collected primarily through self-administered paper-and-pencil questionnaires. At the time of this report, two of these three rounds have been completed. First, students completed a baseline survey in December 2010 and January 2011, shortly before the start of the *HealthTeacher* curriculum in the treatment group schools. Second, students completed a first follow-up survey in October and November 2011, about six months after the end of the curriculum. The final round, a second follow-up survey, is planned for May–June 2012, about one year after the end of the curriculum.

All three surveys are to be administered during the school day by independent field staff hired by Mathematica, though a small fraction of students will be interviewed by phone if they are absent or otherwise unable to complete the survey during school hours. Response rates on the two completed surveys were high; about 95 percent of students in the study sample completed both the baseline and the six-month follow-up surveys. Response rates on the second follow-up survey are also anticipated to be high, in large part because it will take place when students are still in middle school. Students are compensated for their time on the two follow-up surveys; those who completed a follow-up survey in school receive $10 and those who complete it by telephone receive $25.

Outcome Measures

Drawing on the survey data, the study team will construct, and estimate impacts for, a range of outcome measures (Table IV.2). These measures fall into two broad types: (1) sexual risk outcomes, which include both measures of sexual behaviors and their consequences, most notably pregnancy; and (2) intermediate outcomes, which correspond to the mediating factors through which the program would most likely impact behavior (see Figure IV.1 above).

Two of the sexual risk outcomes shown in Table IV.2 will serve as "confirmatory" measures of program impacts: (1) prevalence of sexual intercourse and (2) prevalence of unprotected sexual intercourse.[13] Delaying sexual initiation and increasing the use of birth control methods among sexually active teens are key goals of *HealthTeacher*. Differences in the prevalence of sexual intercourse and unprotected sexual intercourse between members of the treatment and control groups will indicate whether the program was successful in achieving these goals. These primary outcome measures will be evaluated at both follow-ups, though impacts on behavior are more likely to emerge at the long-term, or 12-month, follow-up. While the ultimate goal of *HealthTeacher* is to reduce teen pregnancy, any impacts of *HealthTeacher* on this outcome are unlikely to emerge within the 18-month follow-up period and so it cannot serve as a confirmatory outcome for the study.

Beyond this narrow, confirmatory impact analysis, a comprehensive exploratory analysis will estimate and assess impacts across the range of other outcomes shown in Table IV.2, including both the remaining sexual risk outcomes (shown in the left column) and the full set of intermediate outcomes (shown in the right column). Although these outcomes will not be the basis for confirmatory evidence on the effects of the program, they will provide a complete picture of potential effects of the program on teen sexual risk behaviors and their consequences, and pathways through which the program may affect sexual behaviors. In particular, findings from the analysis of mediating factors will help us to understand the sources and nature of any confirmatory evidence of program impacts on unprotected sex and pregnancy. Examples of important mediators include: students' knowledge and attitudes toward sexual activity and contraceptives, intention to abstain from sexual activity, and refusal skills. For many of these outcomes, we expect changes due to the program to emerge quickly, and therefore may find impacts by the first follow-up.

Table IV.2. Planned Outcomes for Measuring Program Impacts *(HealthTeacher)*

Sexual Risk Outcomes	Intermediate Outcomes
Sexual Behavior:	- Knowledge about contraceptives
- *Prevalence of sexual intercourse*	- Knowledge about STDs
- *Prevalence of unprotected sex*	- Attitudes about sexual activity
- Prevalence of oral sex	- Attitudes about contraceptives
- Prevalence of anal sex	- Intention to engage in sex within a year
- Frequency of sexual activity	- Intention to use contraceptives
- Number of sexual partners	- Refusal skills
Consequences:	
- Pregnancy	
- STDs	

Note: Outcomes shown in italics reflect the confirmatory outcomes for the impact analysis.

[13] To control for multiple hypothesis testing (the increased chance of falsely identifying an impact as statistically significant when examining effects on many outcomes), we are limiting the primary research question to two sexual risk outcomes, which are the focus of stated program objectives.

Sample Size and Statistical Power

To illustrate the study's statistical power for the given sample size and design, in Table IV.3 we show the minimum detectable impacts (MDIs) for the two confirmatory outcomes under different assumptions. As the unit of assignment is schools, the study's statistical power will depend on the level of correlation in the outcomes among students within the same schools (the intraclass correlation coefficient [ICC]). In addition, it will depend on the explanatory power of the regression models used to estimate impacts, given by the R-square. To capture these factors and their uncertainty prior to the analysis, the table displays MDIs for "high" and "low" estimates of both the ICC and R-square.

Table IV.3. Minimum Detectable Impacts for Illustrative Outcomes *(HealthTeacher)*

				Minimum Detectable Impacts	
				R^2 = High Estimate[a]	R^2 = Low Estimate[b]
	ICC	Sample Size (Cluster/Ind.)	Mean Outcome (%)	PPΔ	PPΔ
Second Follow- Up:					
Prevalence of sexual intercourse	Low[c]	17/1,241	25	5.2	5.8
	High[d]	17/1,241	25	9.2	10.3
Prevalence of unprotected sex	Low	17/1,241	5	2.3	2.6
	High	17/1,241	5	4.1	4.5

PPΔ=percentage point change. Minimum detectable impacts are estimated for a 5 percent significance level using a two-tailed test and 80 percent power. The equation used to calculate the minimum detectable impacts for cluster-level random assignment is:

$$MDI = 2.8 \times \sigma \sqrt{\rho \times \left(1 - R_c^2\right) \times \left(\frac{1}{C_T} + \frac{1}{C_C}\right) + (1 - p) \times \left(1 - R_i^2\right) \times \left(\frac{1}{N_T} + \frac{1}{N_C}\right)},$$

where σ is the standard deviation of the outcome measure, ρ is the estimated ICC, C_T and C_C are the number of clusters in the treatment and control groups, and N_T and N_C are the number of individuals in the treatment and control groups. The estimated sample size is based on an 85 percent retention rate for the second follow-up.

[a]High estimate of R-squared is 0.3.

[b]Low estimate of R-squared is 0.1.

[c]Low estimate of ICC is 0.01.

[d]High estimate of ICC is 0.035.

As shown in the table, the MDIs for the two confirmatory outcomes vary widely depending on assumptions, particularly with regard to the ICC. For the prevalence of sexual intercourse (with an estimated control group mean of 25 percent) the MDI ranges from a low of roughly 5 percentage points to a high of more than 10 points.[14] For prevalence of unprotected sex (with an estimated

[14] For both outcome measures, an approximate mean for the control group was estimated based on data from the sample of 8th graders in the 2009 Chicago, Illinois, Youth Risk Behavior Survey (YRBS). The mean for the prevalence of sexual intercourse represents the proportion of youth who reported ever having sexual intercourse. The mean for the prevalence of unprotected sex represents the proportion of the full sample who had sexual intercourse in the last three months and did not use a condom during last sexual encounter.

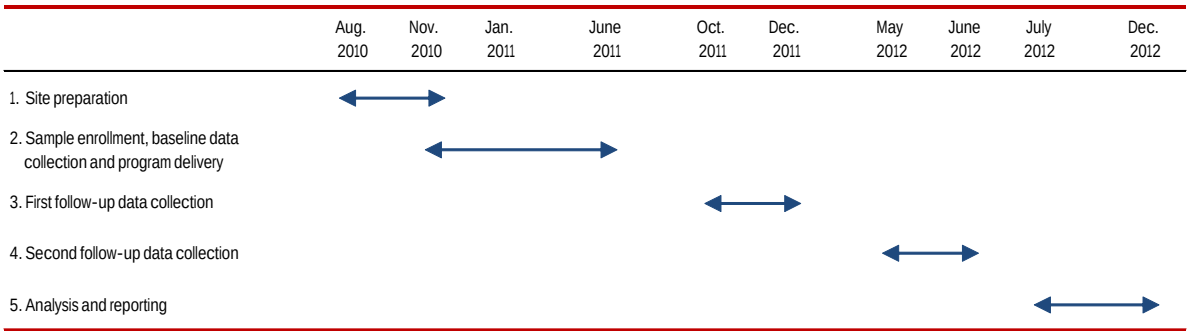
control group mean of 5 percent), the MDI ranges from a low of roughly 2 points to more than 4 points. For both measures, the lower ends of these ranges reflect MDIs at or below the size of impacts seen in prior studies of some school-based sex education programs. However, for the upper ends of these MDIs, only the former measure (prevalence of sexual intercourse) remains within a range seen by prior studies (see, for example, Clark et al. 2005; Coyle et al. 2004).[15] This suggests that the evaluation may only be able to detect relatively large impacts on prevalence of unprotected sex.

Evaluation Timeline and Progress to Date

The evaluation began in 2009, and, by the time of this report, data collection for the second followup survey had been completed. Following enrollment of students in the study sample in late fall 2010, Mathematica administered the baseline survey to the study sample across all 17 schools in December 2010 and January 2011. Subsequently, during the spring 2011 semester, CPS teachers delivered *HealthTeacher* to students in the treatment group schools. In fall 2011—when most students in the study sample had advanced to the 8th grade—the study team administered the first follow-up survey. Administration of the second follow-up survey occurred at the end of this same 8th grade year, in May–June 2012 (see Figure IV.2).

Drawing on these data, the study team will prepare a single comprehensive report detailing the impacts of *HealthTeacher* across the range of outcomes shown above in Table IV.2. This report is planned for completion in December 2012, making it the first of the final impact reports released as part of the overall PPA evaluation.

Figure IV.2. Timeline for the Evaluation of *HealthTeacher*



[15]Given an anticipated control group mean of just 5 percent, the upper bound of 4.5 percentage points on the unprotected sex measure reflects a reduction of 90 percent—an impact that is likely implausible.

V. EVALUATION OF *POWER THROUGH CHOICES 2010*

Teens in foster care are at increased risk of engaging in risky sexual behaviors, becoming pregnant, and having repeat teen pregnancies (Allen et al. 1987). Teen girls in foster care are 2.5 times more likely to become pregnant by the age of 19, and 1.5 times more likely to have a subsequent teen pregnancy, than girls outside of the foster care system (Bilaver and Courtney 2006). For these reasons, youth in foster care and out-of-home placement settings have been identified by HHS as a priority population at high risk for pregnancy. *POWER Through Choices 2010* (*PTC 2010*) is a comprehensive sexual education curriculum focused on pregnancy prevention and designed specifically to meet the unique needs of foster care youth. Mathematica, in partnership with the University of Oklahoma Health Sciences Center (OUHSC) and the Oklahoma Institute for Child Advocacy (OICA), is conducting an experimental evaluation of the effectiveness of *PTC 2010* in reducing pregnancy and associated risk behaviors among foster care youth in California, Maryland, Illinois, and Oklahoma.

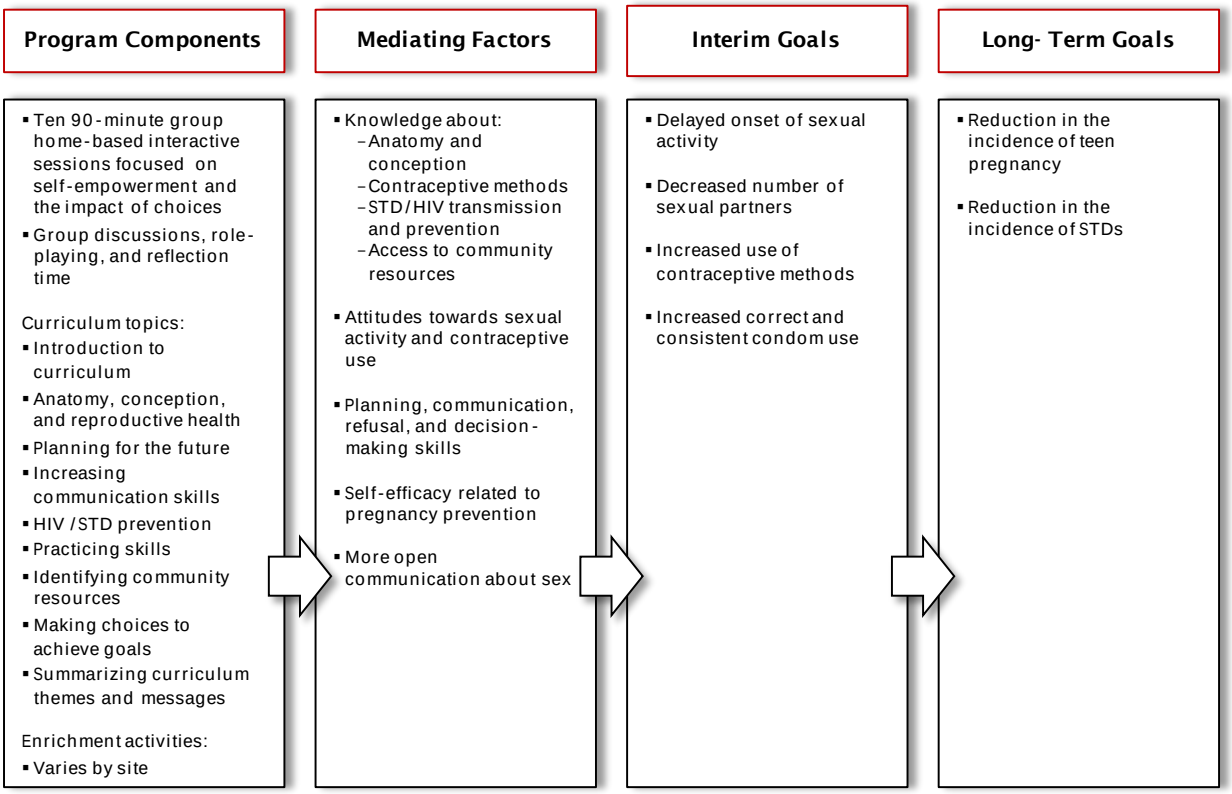
A. Program Features and Evaluation Setting

PTC 2010 is an updated and expanded version of the original *POWER Through Choices* (*PTC*) curriculum, developed in the mid-1990s by the Family Welfare Research Group at the University of California, Berkeley. *PTC* is the only curriculum designed to address specific characteristics of foster care youth that may lead them to engage in risky sexual behaviors, such as a strong need for affection, a desire for a support network, and/or exposure to sexual abuse or violence (OICA and the University of Oklahoma National Resource Center for Youth Services [NRCYS] 2010). It aims to provide foster care youth with information and skills to help them avoid risky sexual behaviors and prevent pregnancy. The 2010 edition of *PTC* includes updated data and new resource information, two new sessions on reproductive health and STDs, and revised role-playing scenarios.[16] The updating was conducted by OICA and NRCYS, under a CDC Promoting Science-based Approaches to Teen Pregnancy Prevention grant.

As with the original PTC curriculum, the logic model for *PTC 2010* (Figure V.1) is grounded in four psychosocial theories—the Health Belief Model, Self-Regulation Theory, Theory of Reasoned Action, and Social and Cognitive Learning Theory. Drawing on these theories, the *PTC 2010* curriculum focuses on (1) increasing participants' understanding of pregnancy and STD risks and their perceived ability to reduce these risks, (2) goal-setting and planning to achieve future goals and avoid pregnancy, (3) the impact of behavioral choices on achieving future goals, and (4) providing opportunities for practicing positive behaviors. Through its skill-building and self-empowerment approach, the curriculum seeks to enable participants to make healthy sexual decisions, build contraceptive knowledge and skills, develop effective communications skills, and learn how to access available resources (OICA and NRCYS 2010).

[16] One session from the original PTC curriculum, "Parents: Making an Informed Choice," was also removed from the curriculum during the updating process. In addition to changes to the curriculum, the updating activities included the development of a revised training-of-trainers manual, fidelity monitoring tool, outcome evaluation tools, and a Direct Care Staff Handbook to support program orientation among group home staff.

Figure V.1. Logic Model for *POWER Through Choices 2010*



The *PTC 2010* curriculum is delivered to small groups of youth in ten 90-minute sessions held ideally twice per week for five weeks The sessions are led by a team of two trained facilitators. The content and objectives of the 10 sessions are the following:

1. **Introduction to *PTC 2010*.** Introduce curriculum, assess participants' knowledge regarding pregnancy prevention and sex education, and demonstrate role playing.
2. **Adolescent Reproductive Health Basics**. Increase knowledge of male and female reproductive anatomy, the process of fertilization and conception, and the menstrual cycle.
3. **Creating the Future You Want**. Identify planning involved in practicing positive sexual behaviors, outline individual choices involved in sexual decision making, and discuss abstinence as a viable choice.
4. **Making Choices Clear**. Help participants to build assertiveness and communication skills related to sexual activity.
5. **Understanding STIs and HIV and How to Reduce Your Risk**. Increase knowledge and understanding of STI/HIV transmission and prevention.
6. **Increasing Contraceptive Knowledge**. Increase knowledge about contraceptive methods.
7. **Practice Makes Perfect**. Discuss level of risk associated with various sexual behaviors, use role playing to demonstrate the importance of dual methods, and practice using a condom.

8. **Using Resources to Support Your Choices**. Discuss ways to improve communication about contraception with foster parents, guardians, and group home staff members, and how to access local sexual and reproductive health resources.
9. **Making Choices That Fit Your Lifestyle**. Help participants to develop a plan for avoiding unwanted pregnancies and STIs, set short- and long-term goals, and identify choices needed to attain goals.
10. **Plan + Prepare + Practice = POWER**. Reinforce themes and messages of the curriculum.

Each session emphasizes the importance of self-empowerment and the impact of individual choices, and employs a combination of role-playing demonstrations, group discussions, and interactive activities to increase sexual and reproductive health knowledge and decision-making and communication skills. Throughout the program, participants are encouraged to envision their future and recognize the importance of making healthy choices in order to accomplish their goals.

Evaluation Sample and Setting

The evaluation of *PTC 2010* will be conducted in 40 foster care group homes—and include up to 120 cohorts of foster care youth within these homes—across four states: California (Kern and Fresco County), Illinois (Cook County), Oklahoma (statewide), and Maryland (Baltimore, Carroll, Frederick, Howard, Montgomery, Prince Georges, and Washington County). In each state, the program will be administered by a local lead agency: Kern County Superintendent of Schools (California site), Illinois Caucus for Adolescent Health (Illinois site), OICA (Oklahoma site), and Planned Parenthood of Maryland (Maryland site). Under centralized training and oversight from OICA and OUHSC, the lead agency in each site is responsible for identifying and recruiting homes to participate in the study, gathering consent for study youth, administering baseline and follow-up surveys, and delivering the *PTC 2010* intervention to homes randomly assigned to the study's treatment group.

All youth between the ages of 13 and 18 residing in participating group homes will be considered for participation; those who provide program and evaluation consent will constitute the study sample. Foster care group homes will be the unit of random assignment for the impact study. Homes recruited for the evaluation will be eligible for randomization multiple times during the study period as their resident population turns over, and a new cohort of sample members is thus defined. The evaluation is expected to enroll an average of three cohorts of youth from each of the approximately 40 participating group homes, with an average of 9 youth per home per cohort, for a total evaluation sample of approximately 1,080 youth.

B. Evaluation Design

The evaluation will test program impacts using an experimental design and cluster random assignment approach. Foster care homes will be randomly assigned to either a treatment group that receives the *PTC 2010* intervention or a control group. There is no alternative program being offered to the control group as part of this project; the study will thus estimate the impact of participating in *PTC 2010* versus receiving existing services. Sample members in both study groups can access existing services offered to teens in the evaluation sites during the study period, some of which may address risk factors related to teen pregnancy (Table V.1).

Table V.1. Summary of Contrast Between the Treatment and Control Conditions (*PTC 2010*)

	Treatment Group	Control Group
Intervention	*PTC 2010*	No intervention
Existing services	- Existing programs and services provided by participating group foster care homes - Services offered through local health departments or community-based organizations, such as screenings, adolescent pregnancy prevention education, and/or reproductive health counseling - School-based health education classes or youth services programs, which may include reproductive health services and/or counseling	

Because the impact of *PTC 2010* may be influenced by the other types of services summarized in Table V.1, it is important to understand these factors in order to accurately characterize the actual treatment-control contrast and interpret study findings. To gain this understanding, the study includes an in-depth implementation analysis that will, among other objectives, gather information about participation in other services by sample members. Information about service receipt will come from three main sources: (1) questions on the participant follow-up surveys conducted as part of the impact study, (2) focus groups with treatment group members, and (3) semi-structured interviews with key program staff and stakeholders in three of the four evaluation sites. Discussion questions will focus on experiences and engagement with the program, as well as the availability and use of other services and activities in the treatment and control communities.

Sample Enrollment and Random Assignment

Enrollment of the study sample involves two steps: (1) the recruitment of foster care group homes in each site and (2) the identification and enrollment of eligible youth residing in participating group homes. For the recruitment of homes, the lead agency in each state will begin by identifying a list of foster care group homes eligible to participate. Each list will be limited to a certain geographic area—for example, Kern and Fresco countries for the California site and Cook County for the Illinois site. Recruitment from these lists began in late 2011 and will continue on a rolling basis until reaching the study's overall target of approximately 40 group homes or 120 cohorts of foster care youth across the four states. The ultimate number of homes may be lower than the targeted number depending on such factors as the average size of each home and resident turnover rate. Other things being equal, fewer homes may be needed if the resident population is larger or turns over more quickly than expected. The ultimate number of homes recruited is also expected to vary by state depending on the size of the catchment area and the willingness of homes to participate. The first cohort recruited in late 2011 totaled 10 homes across three states.

Recruitment of youth for the study will occur on a rolling basis as new homes are recruited or the resident population of participating homes turns over. As noted previously, all youth between the ages of 13 and 18 in each participating group home will be considered for enrollment. Sample intake for each cohort of youth will begin approximately six weeks prior to random assignment. The lead agency in each site will first obtain a list of all youth residing in the homes. Using this list, the agencies will then seek active consent for the program and evaluation from each individual's legally authorized representative (LAR). The LAR may be a state-authorized case worker or lawyer, the individual's biological parent, or a state agency worker, depending on the site's specific circumstances. Only youth whose LARs grant consent will be included in the study.

Once the consent rate for an eligible home reaches at least 80 percent, the home will become eligible for random assignment. This requirement serves two purposes. First, it helps ensure that the

majority of the consent process is completed before random assignment, minimizing any threat of bias due from differential consent by treatment status. Second, it helps identify and screen out homes with low consent rates prior to random assignment. Homes will be randomly assigned in pairs or strata matched on state (California, Illinois, Maryland, Oklahoma), gender (mixed gender home, female-only home, or male-only home), and size (number of beds). Matching on state is necessary to ensure an even geographical distribution of treatment and control homes both across and within states. Matching on gender and size accounts for the existence of single-sex homes and for potentially large variation in home size within states. All homes within a given pair or stratum will have an equal probability of being assigned to either the treatment or control group. Randomization will be conducted by OUHSC on a rolling basis as new homes are recruited into the study and the population of existing homes turns over.

Measuring Program Impacts

Impacts of *PTC 2010* will be analyzed at three time points—upon completion of the program and at approximately 6 and 12 months after program completion. The post-test and 6-month analysis will focus primarily on program impacts on mediating (or intermediate) outcomes, as described in Figure V.1, while the 12-month analysis will focus on program impacts on sexual risk behaviors and associated consequences.

Given the underlying random assignment design, we can calculate unbiased impact estimates at each time point by simply comparing unadjusted mean outcomes between the treatment and control groups. However, to improve the precision of the estimates, we will use regression models to control for covariates, including baseline measures of outcomes. Regression adjustment will also allow us to account for the strata variables used in conducting random assignment and for the correlation of outcomes among youth in the same group home. All models will include stratification variables as covariates and a cluster-level error term, a cluster "random effect."

The main impact estimates will be based on an intent to treat (ITT) analysis that includes all youth who were randomized, regardless of their participation in the program. This approach will yield an estimate of the program's average impact among foster care youth given the opportunity to participate in *PTC 2010*. The magnitude of the ITT estimates could be affected by incomplete or nonparticipation by treatment group members and/or by participation of control group members in other pregnancy prevention services, making these factors important to consider in interpreting the findings. To identify these factors, program staff will track attendance for members of the program group, and data from all sample members on the receipt of other pregnancy-related services will be collected through the follow-up surveys.

Data Sources

All measures will be based on self-reported data drawn from a baseline survey, completed roughly one week before the program begins in the treatment group homes, and three follow-up surveys, administered upon completion of the 10-session curriculum and at 6 and 12 months after program completion. The baseline surveys will be administered in groups as paper-and-pencil questionnaires. The follow-up surveys will be administered in groups or individually depending on whether the study participants still reside in the participating group home. Individual administration of follow-up surveys will occur either in person or through a web-based version of the paper-and-pencil questionnaire. Survey administration will be conducted by the lead agency in each site under the training and guidance of OUHSC. Monetary incentives will be provided for completion of each

survey—$15 for the baseline survey, $20 for the posttest, $30 for the 6-month follow-up survey, and $30 for the 12-month follow-up survey.

Outcome Measures

Drawing on the survey data, the study team will construct, and estimate impacts for, a range of outcome measures (Table V.2). These measures fall into two broad types: (1) sexual risk outcomes, which include both measures of sexual behavior and consequences of this behavior (pregnancy and STDs); and (2) intermediate outcomes, which correspond to the mediating factors through which the program would most likely impact behavior (see Figure V.1).

Table V.2. Planned Outcomes for Measuring Program Impacts (*PTC 2010)*

Sexual Risk Outcomes	Intermediate Outcomes
Behavior: - *Prevalence of sexual intercourse* - *Prevalence of unprotected sex* - Frequency of sexual activity - Number of sexual partners **Consequences:** - Pregnancy - STDs	- Knowledge about male and female anatomy and conception - Knowledge about and attitudes toward contraceptives - Knowledge about STD/HIV transmission and prevention - Self-efficacy related to pregnancy prevention (ability to access and use contraceptives, planning and communication skills, ability to abstain from having unprotected sex) - Attitudes toward sexual activity - Intention to engage in sex within the next year - Intention to use contraceptives

Note: Outcomes shown in italics reflect the confirmatory outcomes for the impact analysis.

Two of the sexual risk outcomes shown in Table V.2 will serve as "confirmatory" measures of program impacts: (1) prevalence of sexual intercourse and (2) prevalence of unprotected sexual intercourse.[17] Delaying sexual initiation and increasing the use of birth control methods among sexually active teens are key goals of *PTC 2010*. Differences in the prevalence of sexual intercourse and unprotected sexual intercourse between members of the treatment and control groups will indicate whether the program was successful in achieving these goals. Both confirmatory outcome measures will be evaluated at the 6- and 12-month follow-ups, though impacts on sexual initiation are more likely to emerge at the 12-month follow-up (when differences between the treatment and control groups will have had sufficient time to emerge). While the ultimate goal of *PTC 2010* is to reduce the incidence of teen pregnancy and STDs, any impacts on these outcomes are unlikely to materialize within the 12-month follow-up period and so they cannot serve as confirmatory outcomes for the study.[18]

[17] In addition to the relative importance of these two outcomes, the confirmatory analysis is limited to them because of the statistical power loss associated with multiple hypothesis testing (the increased chance of falsely identifying an impact as statistically significant as the number of outcomes increases).

[18] In addition, as discussed in Chapter I, self-reported STDs may be subject to considerable reporting error. We plan to conduct a careful assessment of related measures, such as health service use, to determine whether impacts can be rigorously estimated for this outcome measure.

Beyond this narrow, confirmatory impact analysis, a comprehensive exploratory analysis will estimate and assess impacts across the range of other outcomes shown in Table V.2, including both the remaining sexual risk outcomes (shown in the left column) and the full set of intermediate outcomes (shown in the right column). Although these outcomes will not be the basis for confirmatory evidence on the effects of the program, they will provide a complete picture of potential effects of the program on teen sexual risk behaviors and its consequences, and pathways through which the program may affect sexual behaviors. In particular, findings from an analysis of mediating factors—such as knowledge of sexual risk and attitudes toward sexual behavior—will help us understand the sources and nature of any confirmatory evidence of program impacts on sexual risk outcomes. As part of the exploratory analysis, we will also examine program impacts by key subgroups, such as gender.

Sample Size and Statistical Power

As a means of summarizing the statistical power afforded by our sample design, Table V.3 shows the estimated minimum detectable impacts (MDIs) for the two confirmatory outcomes under different assumptions. As the unit of random assignment is a group home, the study's ability to detect statistically significant impacts will depend on the level of correlation in the outcomes among youth within the same home (the intraclass correlation coefficient [ICC]). To assess the sensitivity of the MDIs to the level of correlation, we provide estimates assuming relatively low (0.01) and high (0.035) ICCs. In addition, for each level of the ICC, we provide an estimate of the MDI based on more and less optimistic assumptions about the explanatory power (R^2) of the regression models used to estimate the impacts.

As seen in the table, the lowest MDI for each outcome arises with assumptions of a low ICC and high R^2, whereas the highest MDI arises based on the opposite assumptions. Specifically, for the prevalence of sexual intercourse (projected to have a control group mean of 50 percent), the study is powered to detect a minimum of 8.2 to 10 percentage points depending on the assumed ICC and R^2.[19] For prevalence of unprotected sex (with an anticipated control group mean of 15 percent), the MDI ranges from 5.9 to 7.1 percentage points. [20] Both of these ranges reflect relatively substantial impacts, though within levels detected for programs focusing on reducing sexual risk behaviors with demonstrated evidence of effectiveness; however, these program did not specifically target foster care youth (Coyle et al. 2006; DiClemente 2009; Jemmott et al. 2005).

[19] Prevalence of sexual intercourse is defined as the proportion of youth who ever had sexual intercourse. An approximate mean for the control group is estimated based on data from the 2009 Youth Risk Behavior Survey (YRBS), and represents the proportion that reported ever having sexual intercourse among Hispanic teens in grades 9-12.

[20] Prevalence of unprotected sex is defined as the proportion of youth who had unprotected sex during the months prior to the follow-up. An approximate mean for the control group is based on data from the 2006–2010 National Survey of Family Growth, and represents the proportion who did not use an effective contraceptive during last intercourse (in the past 12 months) among Hispanic teens ages 15–19. (For this measure, teens who did not report having sex during the past 12 months were coded zero.)

Table V.3. Minimum Detectable Impacts for Illustrative Outcomes (*PTC 2010)*

				Minimum Detectable Impacts	
				R^2 = High Estimate[a]	R^2 = Low Estimate[b]
	ICC	Sample Size (Cluster/Ind.)	Mean Outcome (%)	PPΔ	PPΔ
Second Follow- Up:					
Prevalence of sexual intercourse	Low[c]	120/864	50	8.2	9.3
	High[d]	120/864	50	8.8	10.0
Prevalence of unprotected sex	Low	120/864	15	5.9	6.7
	High	120/864	15	6.3	7.1

PPΔ=percentage point change. Minimum detectable impacts are estimated for a 5 percent significance level using a two-tailed test and 80 percent power. The equation used to calculate the minimum detectable impacts for cluster-level random assignment is:

$$MDI = 2.8 \times \sigma \sqrt{\rho \times \left(1 - R_c^2\right) \times \left(\frac{1}{C_T} + \frac{1}{C_C}\right) + (1 - p) \times \left(1 - R_i^2\right) \times \left(\frac{1}{N_T} + \frac{1}{N_C}\right)},$$

where σ is the standard deviation of the outcome measure, ρ is the estimated ICC, C_T and C_C are the number of clusters in the treatment and control groups, and N_T and N_C are the number of individuals in the treatment and control groups. The estimated sample size is based on an 85 percent retention rate for the second follow-up.

[a]High estimate of R-squared is 0.3.

[b]Low estimate of R-squared is 0.1.

[c]Low estimate of ICC is 0.01.

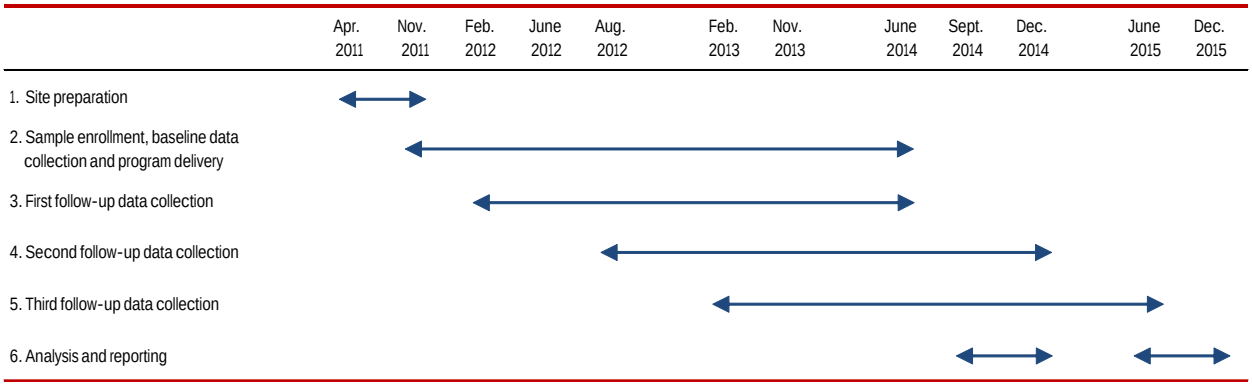
[d]High estimate of ICC is 0.035.

Evaluation Timeline and Progress to Date

The timeline for the *PTC 2010* evaluation extends from start-up planning in 2011 through the submission of a final impact report in 2015 (Figure V.2). Sample enrollment, baseline data collection and program delivery began in late 2011/early 2012 and are expected to continue through June 2014. Post-test data collection will occur between roughly February 2012 and June 2014. The second follow-up data collection, at six months after program completion, will take place between August 2012 and December 2014. The final follow-up data collection will be conducted between February 2013 and June 2015.

A final report presenting the results of the *PTC 2010* evaluation is planned for the end of 2015. A preliminary impact analysis will be conducted based on the post-test and six-month data and reported in an interim report in late 2014.

Figure V.2. Timeline for the Evaluation of *PTC 2010*



THIS PAGE IS INTENTIONALLY BLANK

VI. EVALUATION OF *T.O.P.P.*

Roughly one-quarter of teenage mothers have a second child within 24 months of their first birth, and mothers under age 17 have the greatest risk of a closely spaced repeat pregnancy (Kalmuss and Namerow 1994). Teenage mothers who have rapid repeat pregnancies (within 18 months of the prior birth) are at greater risk of having premature and low birthweight babies and tend to experience poorer educational, economic, and psychological outcomes than mothers who delay subsequent childbearing (Conde-Agudelo et al. 2006, Furstenberg et al. 1987). *Teen Options to Prevent Pregnancy (T.O.P.P.)* is an innovative, clinic-based program that aims to reduce rapid repeat pregnancies among teenage mothers through the provision of telephone-based care coordination and access to mobile family planning services. Mathematica, in collaboration with OhioHealth Research and Innovation Institute and Nationwide Children's Hospital, is conducting an experimental evaluation of *T.O.P.P.*'s effectiveness in reducing repeat teen pregnancies and associated behavioral risks.

A. Program Features and Evaluation Setting

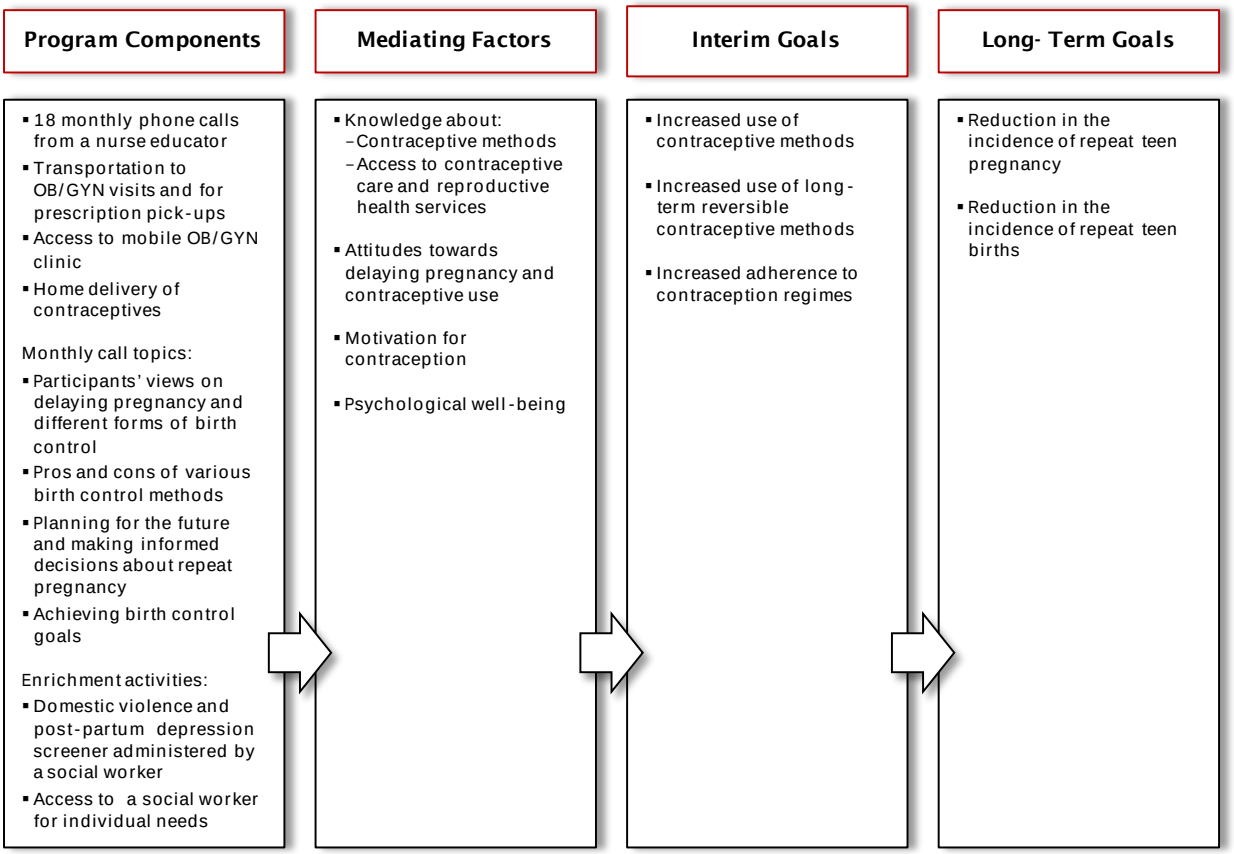
The *T.O.P.P.* program was developed by OhioHealth, an integrated health care system in central Ohio, to reduce rapid repeat pregnancies among teenage mothers served by OhioHealth hospitals and clinics. The program model and underlying theory of change (Figure VI.1) are grounded in the Anderson Behavioral Model of Health Service Use, which posits that increasing perceived need for and access to health services will result in greater health service use and improved health outcomes. The program aims to achieve these goals through two main program components: (1) telephone-based care coordination and (2) access to mobile health services.

For the telephone-based care coordination component of *T.O.P.P.*, program participants receive telephone calls roughly once per month from trained nurse educators. During these monthly calls, the nurse educators use motivational interviewing techniques to educate participants about the importance of delaying subsequent pregnancy, provide information on birth control methods (including abstinence), address any misconceptions inhibiting contraceptive use, and elicit a birth control plan. To facilitate initial and continued contraceptive use, the nurse educators also conduct home visits, as requested, to provide further information on contraceptives, help participants to identify and schedule appointments with providers, and conduct telephone follow-up calls with participants to inquire about side effects or other potential barriers to consistent and continued birth control use.

In addition to monthly calls from a nurse educator, *T.O.P.P.* offers logistical support to help teenage mothers adhere to a contraceptive regimen. This support, managed and overseen by the *T.O.P.P.* program coordinator, is provided in three forms: (1) transportation services, (2) access to a mobile OhioHealth OB/GYN clinic, and (3) a home delivery service. Transportation services are provided by a *T.O.P.P.* program van, which is available to drive participants to and from OB/GYN visits and prescription pick-ups. The mobile clinic, used as part of OhioHealth's *Wellness on Wheels* (*WOW*) program, provides a clinic site to participants who have difficulty accessing stationary clinics. Under *WOW*, the mobile clinic provides prenatal and postpartum services to women ages 12 to 44 in communities with high infant mortality rates, and is stationed once a week at various Columbus high schools. Under *WOW*, the mobile clinic does not provide birth control counseling or administer contraceptives, but does provide patients with a referral for contraceptive services. For *T.O.P.P.*, the *WOW* mobile unit will offer contraceptive services, including the administration of all prescription and long-term reversible contraceptive methods, only to *T.O.P.P.* participants. Up to

once each week, the mobile clinic will be stationed in a library parking lot near four Columbus high schools. *T.O.P.P.* also offers a home delivery service, in which *T.O.P.P.* nurse educators visit the homes of participants to administer contraceptive injections.

Figure VI.1. Logic Model of the *T.O.P.P.* Intervention



The *T.O.P.P.* program will be delivered to each program participant over an 18-month period, a time frame prior research identifies as the minimum suggested interpregnancy interval (Conde-Agudelo et al. 2006). At the start of the program, each participant will be assigned to a nurse educator trained in Motivational Interviewing for the duration of the program. The assigned nurse educator will conduct the monthly motivational interviewing calls and provide related telephone-based care coordination services, including referrals to the *T.O.P.P.* program coordinator for *T.O.P.P.*'s transportation and home delivery services. At the beginning of the program, a *T.O.P.P.* social worker will also administer domestic violence and post-partum depression screeners to each program participant to help identify particular needs and challenges.[21] Over the course of the program, the nurse educators can also refer *T.O.P.P.* participants, as needed, to the social worker for further clinical assessment and/or referral assistance to access a mental health provider or other community resources.

[21] Particular challenges and needs may relate to employment and training, education/school, child care, housing, food and other material assistance, legal services, insurance, mental health care, substance abuse and counseling, and domestic coercion and violence.

Evaluation Sample and Setting

The evaluation will be conducted in the metropolitan Columbus, Ohio, area. The study population comprises eligible, low-income OhioHealth patients who are pregnant or have just given birth. Participants will be recruited from seven OhioHealth women's health clinics and the postpartum units of five OhioHealth hospitals. These facilities serve adolescents from seven central Ohio counties, including Fairfield, Franklin, Delaware, Licking, Madison, Pickaway, and Union. However, all of the participating clinics and hospitals are located in Franklin County, with the exception of one hospital located in Delaware County.

To be eligible for the evaluation, women must be between the ages of 10 and 19, at least 28 weeks pregnant or less than nine weeks postpartum, and enrolled in Medicaid. Due to the telephone-based nature of the intervention, patients must also have regular telephone service and speak English. The evaluation is expected to enroll approximately 600 young women, half of whom will be randomly assigned to the program group and will have the opportunity to participate in *T.O.P.P.* Enrollment of these women in the study will take place on a rolling basis over a two-year period, with those selected for the program group beginning to participate in *T.O.P.P.* soon after study enrollment.

The population targeted for the program largely consists of young women from high-risk backgrounds. Franklin County, where most OhioHealth clinics and hospitals are located, has a teen birth rate higher than both state and national averages. In 2008, the birth rate for teens ages 15 to 17 in Franklin County was 24.9 (per 1,000 females), compared to 19.7 for Ohio and 21.7 for the United States (Nationwide Children's Hospital 2010). The county average, however, masks significant racial and ethnic disparities in Franklin County. In 2008, the birth rates for black (37.8) and Hispanic (69.8) teens ages 15 to 17 were two to four times higher than the rate for non-Hispanic whites (16.0). Roughly one-fifth of all teen births in Columbus, Ohio, are repeat births (Child Trends 2008).

B. Evaluation Design

The evaluation will test program impacts using an experimental design and random assignment of individuals to either a treatment or control condition. The main difference between the treatment and control conditions is that treatment group members will have the opportunity to participate in *T.O.P.P.* There is no equivalent or alternative program offered to members of the control group. However, since the study sample consists of OhioHealth patients, both the treatment and control group will have access to and receive the benefits of existing services provided through the OhioHealth system, such as OhioHealth's current standard of OB/GYN care, the mobile prenatal services provided through *PRIM*, and any other community outreach services. In addition, study participants can access other clinic- or community-based services available to teens in the OhioHealth catchment area (Table VI.1). Thus, the evaluation will compare the outcomes of teen mothers with access to *T.O.P.P.* and other existing services against outcomes of teen mothers with access only to existing services.

Table VI.1. Summary of Contrast Between the Treatment and Control Conditions *(T.O.P.P.)*

	Treatment Group	Control Group
Intervention	*T.O.P.P.*	No intervention
Existing Services	- OhioHealth's five hospital's current standard of OB/GYN care - Prenatal care services provided by OhioHealth to women ages 12 to 44 through the *Wellness on Wheels (WOW)* mobile OB/GYN trailer. The *WOW* unit is stationed once a week at high schools in communities with high infant mortality rates. - Other services and programs available in Columbus, Ohio, including those provided by Planned Parenthood and other community organizations.	

Because the impact of *T.O.P.P.* may be affected by the availability of other programs and services in the study area, it is critical that we understand these factors so that we can accurately characterize the actual treatment-control contrast and interpret the study's findings. To gain this understanding, we plan to conduct an in-depth implementation study that will provide a detailed assessment of program implementation, as well as gather information about participation in other services and activities by members of both the treatment and control groups during the study period. Information about service receipt will come from several sources: (1) questions on the participant follow up surveys conducted as part of the impact study, (2) program data on participation and service use (for treatment group members), (3) semi-structured interviews with program staff and other local health professionals and administrators, and possibly (4) interview data from treatment group members on their perceptions of the program.

Sample Intake and Random Assignment

Sample enrollment began in October 2011 and will continue on a rolling basis for approximately two years, or until roughly October 2013. As noted above, recruitment of sample members will occur in seven OhioHealth women's health clinics and five OhioHealth hospitals. Eligible clinic patients will be identified through weekly queries of OhioHealth's electronic scheduling system, which will produce a list of eligible patients and their next scheduled clinic appointment. In hospitals, OhioHealth's inpatient registration system will be used to identify eligible patients who have just given birth.

After eligible patients have been identified, sample enrollment will follow a three-step process at each recruitment location. First, eligible patients will be approached by an OhioHealth care provider during their regularly scheduled OB/GYN clinic appointment or in the postpartum unit of the hospital and told about the opportunity to participate in a program that is part of a study being conducted by OhioHealth. Second, among patients who agree to learn more about the program and study, *T.O.P.P.* program and local evaluation staff will follow up with information about the study and the potential opportunity to participate in the *T.O.P.P.* program, should they provide consent and be assigned to the treatment group. In most cases, this follow-up will occur while the patient is still at the OhioHealth clinic or hospital. However, in some cases, the initial recruitment visit will take place in patients' homes.

The third step in the enrollment process involves providing consent and completing a baseline survey. All study participants must provide consent for the program and study. For participants under age 18, consent must be provided by their parents or legal guardians. Consent forms will be distributed to eligible participants at the time of initial recruitment, but may be collected at a later date. In addition to providing active consent, women must complete the baseline survey in order to be eligible for the study and random assignment. In most cases, baseline data will be collected at the

same time as consent, either at an OhioHealth site or at an off-site location. Eligible patients who provide consent but do not complete the baseline survey will not be randomly assigned (and will not be part of the evaluation sample).

Upon completion of the baseline survey, OhioHealth staff will use Mathematica's web-based sample management system (SMS) to record information on each sample member. Once this information is entered into the SMS, random assignment will follow on an individual, rolling basis. Following this rolling process will enable *T.O.P.P.* program staff to notify study participants of their program or control status soon after study enrollment, thereby allowing nurse educators to contact women selected for the program group quickly.

Random assignment will be based on a random allocation sequence, known as permuted block randomization, with stratification by recruitment site and age. Under this method, a random sequence of treatment and control assignments is generated in advance of sample enrollment within each stratum, with the sequence constrained to achieve a roughly constant allocation of treatment ("T") and control ("C") assignments throughout the enrollment period. To achieve this constant allocation, the assignment sequence is generated by first creating smaller "blocks" of assignments of varying lengths (for example, "TC" or "TTCC"), which are then combined to create the full randomization sequence. For this study, we will use a 1:1 allocation ratio to ensure a balanced sample, and employ random block sizes up to six characters. Separate randomization lists will be created for every combination of strata levels. With 12 recruitment locations (7 OhioHealth clinics and 5 OhioHealth hospitals) and two age groups (those who are under 18 years old and those who are 18 or 19 years old), the study will employ 28 strata. All staff involved in recruitment will be blinded to the randomization sequence for each stratum. Assignment results will be revealed through the web-based SMS as each participant is enrolled in the study.

Measuring Program Impacts

We will analyze impacts using data collected approximately 6, 18, and 30 months after study participants complete the baseline survey. Because *T.O.P.P.* is an 18-month program, the 18-month analysis will occur at program completion and the 30-month analysis will occur one year after program completion. Data for the 6-month survey will be collected while the program is ongoing and will be used to measure interim program impacts.

Given the underlying random assignment design, we will be able to calculate unbiased impact estimates at each time point by simply comparing unadjusted mean outcomes between the treatment and control groups. However, to improve the precision of the estimates, we will use regression models to control for covariates, including baseline measures of outcomes. Regression adjustment will also allow us to account for the strata variables used in conducting random assignment, which will be included as covariates in the regression models.

The main impact estimates will be based on an intent-to-treat (ITT) analysis, in which all sample members are included in the analysis regardless of their participation in the program. This approach will yield an estimate of the program's average impact among young women given the opportunity to participate in *T.O.P.P.*, which is the most relevant estimate for a voluntary program. The magnitude of the ITT estimates could be affected by incomplete or nonparticipation by treatment group members and/or by participation of control group members in other pregnancy prevention services, making these factors important to consider in interpreting the findings. To identify these factors, program staff will track data on participation and service use for members of the treatment

group for use in the impact analysis. In addition, data from all sample members on the receipt of other pregnancy-related services will be collected through the follow-up surveys.

Data Sources

Data will be drawn from the baseline survey and the three follow-up surveys administered 6, 18, and 30 months later. The baseline survey will be a paper-and-pencil questionnaire administered individually after consent by recruiters for the evaluation. Follow-up data will be collected through telephone interviews conducted by trained Mathematica interviewers. Incentives will be provided for survey completion, in the amounts of $10 for the baseline and 6-month surveys, $25 for the 18-month survey, and $50 for the 30-month survey. All incentives will be paid in the form of gift cards to local stores.

The survey data will be supplemented with administrative data from OhioHealth medical records and Ohio state birth certificates. The medical records will provide information for both treatment and control group members of service receipt from OhioHealth women's health clinics and hospitals. The state birth certificates will provide information on subsequent births to study participants. Both administrative data sources will be merged with the survey data for use in the impact analysis.

Outcome Measures

The study team will construct and estimate program impacts on a range of outcome measures (Table VI.2). These include sexual risk outcomes, such as pregnancy, STDs, and associated sexual risk behaviors, and intermediate outcomes—such as attitudes, knowledge, and intentions—that correspond to the mediating factors through which the program, if successful, would most likely impact sexual behaviors and their consequences (see Figure VI.1).

Two of the sexual risk outcomes shown in Table VI.2 will serve as primary, or "confirmatory," measures of program impacts: (1) prevalence of unprotected sexual intercourse and (2) incidence of repeat pregnancy. The ultimate goal of the program is to reduce repeat teen pregnancies, and consistent use of a birth control method is the key behavioral pathway targeted by *T.O.P.P.* to achieve this goal. Prevalence of unprotected sex will be measured by the proportion of the sample that had sexual intercourse without the use of any (effective) contraceptive method. Reflecting the program's goal of reducing repeat teen pregnancies that occur within 18 months of a prior birth, we will measure the incidence of repeat pregnancy as the proportion of sample members who report a pregnancy between baseline and the 18 month follow-up survey. To measure the potential for sustained program impacts beyond this 18-month period, we will also estimate impacts on the incidence of repeat pregnancy at the 30-month follow-up survey. To address the possible bias resulting from survey nonresponse or misreporting of pregnancy history, we will also estimate impacts on the incidence of subsequent births as measured by the Ohio state birth certificates.

Beyond the more narrow confirmatory impact analysis, a comprehensive exploratory analysis will estimate and assess impacts across the range of other outcomes shown in Table VI.2, including the remaining sexual risk outcomes (shown in the left column) and the full set of intermediate outcomes (shown in the right column). Although these outcomes will not be the basis for confirmatory evidence on the effects of the program, they will provide a complete picture of potential effects of the program on teen sexual risk behaviors and their consequences, and pathways through which the program may affect sexual behaviors. In particular, findings from the analysis of intermediate outcomes, or mediating factors, will help us to understand the sources and nature of

any confirmatory evidence of program impacts on the prevalence on unprotected sex and repeat pregnancy. Exploratory analyses will also be conducted to understand whether and how program impacts vary by key subgroups, such as by gender and/or race/ethnicity.

Table VI.2. Planned Outcomes for Measuring Program Impacts *(T.O.P.P.)*

Sexual Risk Outcomes	Intermediate Outcomes
Sexual Behavior: - *Prevalence of unprotected sex* - Prevalence of sexual activity - Frequency of sexual activity - Number of sexual partners **Consequences:** - *Repeat pregnancy* - Repeat births - Interpregnancy interval - STDs	- Attitudes toward delaying pregnancy and contraceptive use - Knowledge about birth control methods - Intention to use contraception - Access to family planning and reproductive health services

Note: Outcomes shown in italics reflect the confirmatory outcomes for the impact analysis.

Sample Size and Statistical Power

To illustrate the statistical power of the evaluation, in Table VI.3 we show the estimated minimum detectable impacts (MDIs) for the two confirmatory outcomes. To assess the sensitivity of the MDIs to the explanatory power (R^2) of the impact regression models, we provide estimates of the MDI for relatively low and high values of R^2. The MDI for the prevalence of unprotected sexual intercourse (evaluated at an estimated mean of 20 percent), is 9.4 percentage points assuming an R^2 of 0.1. [22] Under the more optimistic assumption of an R^2 of 0.3, the study is powered to detect a minimum of 8.3 percentage points difference. For the incidence of repeat pregnancy, at an estimated mean of 45 percent, the study is powered to detect a minimum impact of 10.6 to 12 percentage points, depending on the value of R^2.[23] The lack of rigorous evaluations of individualized interventions in a clinic-based setting targeting sexually active or pregnant teens makes it difficult to assess how the expected impacts of the *T.O.P.P.* program compare to the size of the estimated MDIs. However, the size of the estimate for unprotected sexual intercourse is within the range detected by programs focusing on reducing sexual risk behavior with demonstrated evidence of effectiveness (Coyle et al. 2006; DiClemente 2009; Jemmott et al. 2005).

[22] Prevalence of unprotected sex is defined as the proportion of youth who had unprotected sex during the months prior to the follow-up. An approximate mean for the control group is based on data from the 2006-2010 National Survey of Family Growth, and represents the proportion who did not use an effective contraceptive during last intercourse (in the past 12 months) among youth ages 15-21 who had sex in the past 12 months. (For this measure, teens who did not report having sex during the past are excluded, as the target population for the evaluation is pregnant or parenting teens.)

[23] The incidence of repeat pregnancy is defined as the proportion of youth who had at least one pregnancy between baseline and the 30-month follow-up. An approximate mean for the control group was estimated based on birth files for 2009 published by the Centers for Disease Control and Prevention (CDC). The estimate represents the proportion of births that were second or higher order, among youth ages 13-22 residing in Franklin County, Ohio.

Table VI.3. Minimum Detectable Impacts for Illustrative Outcomes *(T.O.P.P.)*

	Sample Size	Mean Outcome (%)	Minimum Detectable Impacts: R^2 = High Estimate[a] PPΔ	Minimum Detectable Impacts: R^2 = Low Estimate[b] PPΔ
First Follow- Up:				
Prevalence of unprotected sexual intercourse	510	20	8.3	9.4
Second Follow- Up:				
Incidence of repeat pregnancy	480	45	10.6	12.0

PPΔ=percentage point change. Minimum detectable impacts are estimated for a 5 percent significance level using a two-tailed test and 80 percent power. The equation used to calculate the minimum detectable impacts for individual-level random assignment is:

$$MDI = 2.8 \times \sigma \sqrt{\rho \times \left(1 - R_i^2\right) \times \left(\frac{1}{N_T} + \frac{1}{N_C}\right)},$$

where σ is the standard deviation of the outcome measure, and N_T and N_C are the number of individuals in the treatment and control groups. The estimated sample size is based on an 85 percent and 80 percent retention rate for the first and second follow-up, respectively.

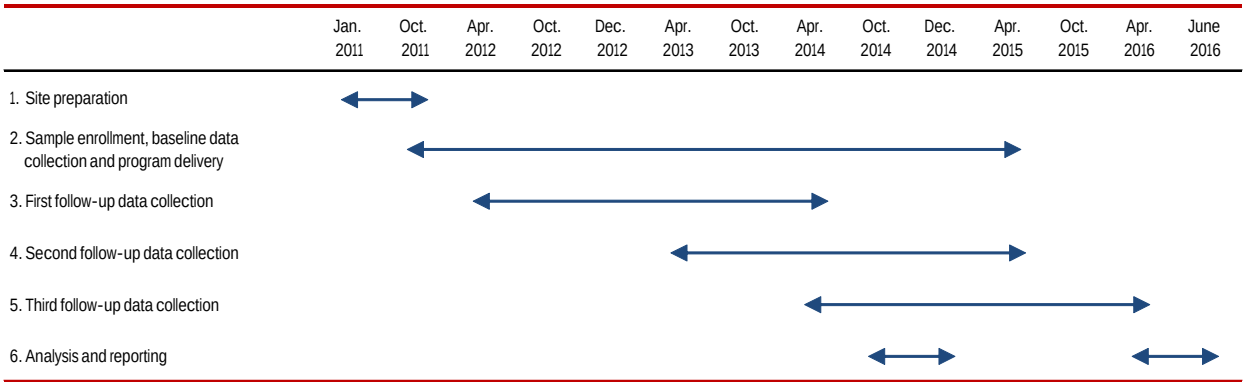
[a]High estimate of R-squared is 0.3.

[b]Low estimate of R-squared is 0.1.

Evaluation Timeline and Progress to Date

The timeline for the *T.O.P.P.* evaluation begins with start-up planning in early 2011 and continues through submission of a final impact report in 2016 (Figure VI.2). Sample enrollment, baseline data collection and program delivery started in October 2011 and will continue through approximately April 2015. Data collection for the first follow-up will occur between approximately April 2012 and April 2014, and for the second follow-up between approximately April 2013 and April 2015. Data collection for the third follow-up is expected to be completed by April 2016. A final impact report presenting the results of the *T.O.P.P.* evaluation is planned for mid-2016. Preliminary impact analysis will be conducted after completion of the six-month follow-up for the full sample, and reported in an interim report in late 2014.

Figure VI.2. Timeline for the Evaluation of *T.O.P.P.*



VII. EVALUATION OF *TEEN PEP*

Teenagers can be particularly susceptible to external influences, especially the opinions and behaviors of their friends and classmates. To harness the power of peer influence for positive change, the Princeton Center for Leadership Training (PCLT), in collaboration with HiTOPS, developed the *Teen Prevention Education Program* (*Teen PEP*), a school-based peer-to-peer comprehensive sexual education program aimed at reducing risky sexual behaviors and teen pregnancy. *Teen PEP* trains 11th and 12th grade students (peer educators) to conduct skills-based sexual health workshops with younger students. During the program planning and implementation stages, *Teen PEP* engages a variety of stakeholders at participating schools and provides resources and strategies to foster a school-wide culture of positive peer pressure that promotes responsible and healthy sexual decision making among teens.

Since its inception in 1995, *Teen PEP* has grown to include 50 schools across New Jersey, as part of a collaboration with the New Jersey Department of Health and Senior Services.[24] In 2006, the program expanded to North Carolina, where it has been implemented to date in six schools. While the program in New Jersey (New Jersey *Teen PEP*) has been previously evaluated, prior studies of the program's impact on sexual behaviors have not met HHS evidence standards for study quality and rigor (Goesling and Trenholm 2010).[25] Mathematica, in collaboration with PCLT, is conducting the first rigorous, experimental evaluation of *Teen PEP*, drawing on a combined sample of 16 high schools across New Jersey and North Carolina. (Mathematica and PCLT are partnering with Abt Associates Inc. on the study in the North Carolina schools.[26])

A. Program Features and Evaluation Setting

The *Teen PEP* logic model (Figure VII.1) is grounded in social learning theory, the health belief model, and principles of youth development. Social learning theory posits that individuals learn by observing the behaviors of others and the consequences of those behaviors (Ormond 1999). The health belief model contends that greater awareness of the consequences of risky health behaviors, knowledge of risk mitigation strategies, and self-confidence in one's ability to avoid a negative health outcome will lead to positive behavioral change (Glanz et al. 2002). Drawing on tenets of these models, *Teen PEP* employs a multidimensional, youth development-oriented approach that includes comprehensive sexuality education, a broad sexual health focus, peer education and mentoring, and a school-wide sexual health campaign to promote broader cultural change within the school community.

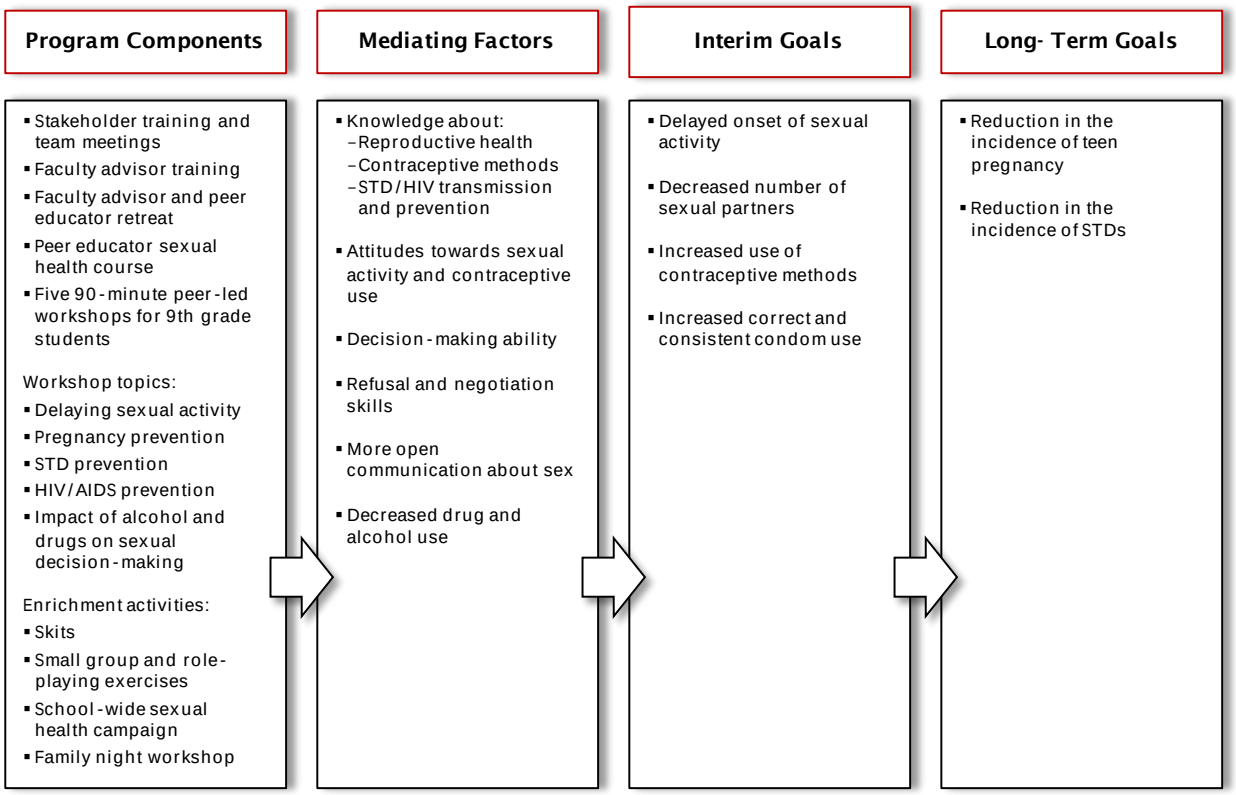
---

[24] Additional sponsors of New Jersey *Teen PEP* include the New Jersey Department of Education, the New Jersey Department of Human Services, and the Horizon Foundation of New Jersey.

[25] The HHS evidence standards can be found at: http://www.hhs.gov/ash/oah/oah-initiatives/tpp/eb-programs-review.pdf, accessed February 15, 2012. The study ratings for teen pregnancy prevention studies reviewed to date using HHS evidence standards can be found at: http://www.hhs.gov/ash/oah/oah-initiatives/tpp/PPRER_Studies_Table_formatted_Final.pdf, accessed February 15, 2012.

[26] As noted in Chapter I, PCLT is receiving TPP Tier 2 grant funding to implement the *Teen PEP* program in North Carolina. The TPP grants require and include funding for an external evaluation of the program by a local evaluator, in this case, Abt Associates Inc.

Figure VII.1. Logic Model of the *Teen PEP* Intervention



Teen PEP is implemented in two phases. During the first (planning) phase, which typically covers a 6- to 12-month period, PCLT works collaboratively with school and district administrators and staff to lay the foundation for successful long-term implementation. Liaisons in each participating school assemble a stakeholder team—consisting of school faculty and administrators, community members, parents, and students—to oversee the planning and implementation process. Key planning tasks during the first program phase include the selection of two faculty advisors to coordinate the program and teach the *Teen PEP* course to peer educators, the training of the faculty advisors by PCLT and HiTOPS, and the selection of 15 to 24 10th and/or 11th grade students to serve as peer educators the following school year (when they are juniors and seniors). The selection of peer educators is based on an application, faculty recommendations, and an individual and group interview. Just prior to the second program phase, the peer educators and faculty advisors participate in a three-day training retreat to prepare them for program implementation.

The second (implementation) phase involves two key components: (1) delivery of the *Teen PEP* sexual health class for 11th and 12th grade peer educators and (2) delivery of sexual health workshops by the peer educators to underclassmen, typically in the 9th grade. The *Teen PEP* sexual health class, team-taught by the two trained faculty advisors, is a 10-unit, credit-bearing course (usually delivered during health or physical education class sessions) that prepares peer educators for their role as workshop facilitators and sexual health advocates. While enrolled in the sexual health class, the peer educators are required to conduct at least five 90-minute sexual health workshops with 9th graders, as well as one outreach workshop with parents of 9th graders.

The PPA impact evaluation will test the effects of *Teen PEP* on 9th grade students in participating high schools. The peer-led workshops for 9th graders will cover the following topics:

(1) postponing sexual involvement, (2) pregnancy prevention and contraceptives, (3) STD prevention, (4) HIV/AIDS prevention, and (5) the impact of alcohol and other drugs on sexual decision making. Each of these workshops open with skits to introduce the relevant sexual health issue. Peer educators then give an informational presentation on related sexual health topics, such as abstinence, correct and consistent use of condoms, hormonal birth control methods, and STD transmission and prevention. Finally, small group discussions and role play activities are used to help students build communication, problem-solving, decision-making, and self-management skills to support sexual decision making. The family night workshop focuses on opening communication between parents and children about sex and is structured similarly, featuring skits, an informational presentation, and group activities.

In addition to the workshops for 9th graders and their parents, peer educators coordinate—in collaboration with the faculty advisors, school administrators, and other faculty—a school-wide campaign that aims to reform school culture regarding sexual activity, alcohol, and drugs. Through the workshop and cultural change components of the program, *Teen PEP* seeks to reduce sexual risk behaviors and ultimately rates of pregnancy and STDs among high school students.

Evaluation Sample and Setting

The *Teen PEP* evaluation will be conducted in 16 high schools in New Jersey and North Carolina. Once recruited, each of the 16 schools will be randomized to either an "early start" (treatment) group or a "late start" (control) group. At each treatment school, the *Teen PEP* workshops will be administered to eligible 9th grade students. In New Jersey, program administration will take place throughout the school year; in North Carolina, it will take place during the spring semester in order to accommodate the block scheduling used by high schools in that state.[27] Control schools will be eligible to implement *Teen PEP* after completion of the study, including all data collection from the study sample, roughly three years after enrollment in the study. During the study period, control schools will be provided with materials and a one-day training to support implementation of *Safe Dates*, an adolescent dating abuse prevention program.[28] Implementation of *Safe Dates* in the control schools is optional and will be directed toward juniors and seniors to avoid "contamination" of the control group with services aimed at some of the same mediating factors as *Teen PEP*, such as decision-making skills.

To date, 12 schools have been recruited for the *Teen PEP* study. This includes four schools in New Jersey, which will begin study implementation—including sample enrollment, baseline data collection, and programming (if appropriate)—in the upcoming 2012–2013 school year; and it includes eight schools in North Carolina, four that began study implementation during the 2011–2012 school year (cohort 1) and four that will begin implementation in 2012–2013 (cohort 2). In

[27] North Carolina schools use block scheduling (instead of the tradition year-long course scheduling used in New Jersey). On a block schedule, high school students attend four 90-minute classes each day rather than six 50-minute classes each day. Because of the increased time in each course, a traditional school year's worth of study is completed in one semester.

[28] See http://www.hazelden.org/web/go/safedates, accessed on February 16, 2012, for more information on the Safe Dates curriculum. Control schools receive a *Safe Dates* implementation manual and the accompanying CD-ROM. No financial support is provided to schools for the implementation of *Safe Dates*.

both states, recruitment for the study sample continues with the aim of all schools beginning program implementation by the 2012–2013 school year. However, should sample recruitment through this second cohort fall short of the targeted 16 study schools, recruitment will be extended in both states into a third cohort of schools that will begin study implementation during the 2013–2014 school year. Consistent with the goals of PPA, all 12 schools recruited for the study to date have large numbers of lower-income students as well as a diversity of racial-ethnic groups, including large numbers of African American and Hispanic students. The remaining schools recruited for the study are expected to have similar socio-demographic characteristics.

The evaluation sample will consist of eligible 9th grade students in participating high schools. In most schools, all students in the 9th grade at the time of sample enrollment will be eligible to participate in the *Teen PEP* workshops and evaluation, and will be enrolled in the study after obtaining parental consent for the evaluation. However, in very large high schools, it may not be financially or logistically feasible to offer the workshops to all 9th graders, in which case sample enrollment will be limited to a subgroup of students who will eligible for the program (if the school is selected for the treatment group).[29] Regardless, we expect that roughly 200 students will be targeted for sample enrollment in each participating school, yielding an approximate sample of 3,200 students for the evaluation.

B. Evaluation Design

The evaluation will test program impacts using an experimental design and cluster random assignment approach. The main difference between the treatment and control conditions is that the treatment schools will deliver the *Teen PEP* workshops during the school year following random assignment, whereas the control group schools will not begin implementing *Teen PEP* until after the PPA impact study has been completed. During the study period, schools in the control group will continue to provide existing programming, such as health and physical education classes, and will have the option of administering *Safe Dates* to juniors and seniors. Thus, the evaluation will compare the outcomes of 9th graders with access to *Teen PEP* workshops (and any existing services in the community) to the outcomes of 9th graders with access only to existing services.

A range of community- and school-based sexual and reproductive health services are available to New Jersey and North Carolina high school students, as shown in Table VII.1. In both states, schools are required to offer some form of comprehensive sexual education and, in each of the study schools, students could eventually participate in workshops or assemblies focused on STDs or pregnancy prevention (in addition to what might be available through *Teen PEP).* This is particularly true of New Jersey, where organizations like Planned Parenthood offer school-wide assemblies and other "modest" programming to schools throughout the state.

[29]To date, only one school, in North Carolina, has limited sample enrollment; due to the timing of its health and physical education classes, only half the 9th grade students will be eligible to participate in the program should the school be selected.

Table VII.1. Summary of Contrast Between the Treatment and Control Conditions *(Teen PEP)*

	Treatment Group	Control Group
Intervention	*Teen PEP*	No intervention
Existing Services in NJ	- Services offered through school-based youth services programs such as obstetric/gynecological exams, STD testing, family counseling, drug and alcohol abuse prevention, and family life education - Dating violence and rape prevention programs - Annual school-wide assemblies or regular structured workshops led by Planned Parenthood, HiTOPS, or other community-based organizations on sexual health - Peer mentoring programs	
Existing Services in NC	- Services offered through the local health departments such as screenings, adolescent pregnancy prevention education, and counseling. - School-based health education classes, including some discussion on reproductive health, abstinence, and relationships and dating	

Because the impact of *Teen PEP* may be influenced by the other types of programming and services summarized in Table VII.1, it is critical that we understand these factors so we can accurately characterize the actual treatment-control contrast and interpret the study's findings. To gain this understanding, we plan to conduct an in-depth implementation study that will provide a detailed assessment of program implementation, as well as gather information about participation in other services by treatment and control group members through the follow-up surveys conducted as part of the impact study. Information about service receipt will come from three main sources: (1) questions on the participant follow up surveys, (2) focus groups with treatment group members, and (3) semi-structured interviews with staff from and stakeholders affiliated with both the treatment and control group schools. Discussion questions will focus on experiences and engagement with the program (from the perspective of treatment group members and schools), as well as the availability and use of other services and activities in the treatment and control school communities.

Sample Enrollment and Random Assignment

Sample enrollment follows a two-stage process. The first stage involves the recruitment of schools for the study, and entails several steps, including: (1) initial discussions between PCLT and district- and school-level leadership about *Teen PEP* and the PPA evaluation; (2) the collection of data from interested schools on the student population and existing sexual and reproductive health services, to determine the school's suitability for the evaluation; and, finally, (3) in-depth discussions between PCLT, Mathematica staff, and representatives of eligible schools about the program and evaluation. To participate in the study, each school must complete the *Teen PEP* application and sign an agreement that commits the school and district to adhere to the requirements of study participation.

The second phase of the sample enrollment process involves the collection of active consent from parents of eligible 9th graders in each study school. Approximately two months prior to the delivery of the peer-led workshops to 9th graders, data collectors trained by Mathematica will begin distributing consent forms to each school and scheduling times for collection of those forms from school administrators. The resulting study sample will consist of eligible, consented 9th grade students at the 16 schools participating in the evaluation. As noted above, approximately 200 students per school will be recruited and enrolled into the study sample.

Because implementation of *Teen PEP* requires an intensive planning phase that begins prior to the start of the school year, random assignment of schools must occur before we begin gathering consent for the eligible 9th grade students. Neither eligible students nor their parents will be informed of the school's treatment status until after consent gathering and baseline survey administration have been completed, ensuring the integrity of the experimental design.

Schools will be assigned to either the treatment or control condition using a stratified random assignment approach. This approach, which can increase the precision of impact estimates and further ensure balanced study groups, has up to two stages. First, schools will be stratified by state and year of sample enrollment. Second, should the schools within a given state/year stratum display notable differences in expected rates of sexual risk behavior, these schools may be further divided into sub-strata. Data to form these possible sub-strata include a variety of factors that, together, may be associated with sexual risk outcomes, including: (1) school setting (urban, suburban, rural); (2) racial/ethnic distribution of the school; (3) socioeconomic composition of the school (such as percentage of students receiving free or reduced-price lunch); and (4) academic performance (such as average SAT score and/or graduation rate).

Random assignment has been conducted for the 12 schools recruited to date across four strata. This includes two strata of North Carolina schools, one containing the four schools recruited for study implementation in the 2011–2012 school year and the other containing the four schools recruited for implementation in the 2012–2013 school year. It also includes two strata of New Jersey schools, each containing a pair of the four New Jersey schools recruited for implementation in the 2012–2013 school year. All assignments were made using a random number generator (Rand function).

Measuring Program Impacts

Impacts of *Teen PEP* will be analyzed approximately 6 and 18 months after program completion. All outcomes will be evaluated at both follow-up time points. However, analysis based on the 6-month follow-up will focus more on mediating outcomes, while the analysis of the longer, 18-month follow-up will focus most on sexual risk outcomes. Data from participants at all study schools will be pooled for all impact analyses, as the study is not designed and powered to detect differential program impacts by state or school.

Given the underlying random assignment design, we can calculate unbiased impact estimates at each time point by simply comparing unadjusted mean outcomes between the treatment and control groups. However, to improve the precision of the estimates, we will use regression models to control for covariates, including baseline measures of outcomes. Regression adjustment will also allow us to account for the strata variables used in conducting random assignment and for the correlation of outcomes among students in the same school. All models will include stratification variables as covariates and a cluster-level error term, a cluster "random effect."

The main impact estimates will be based on an intent-to-treat (ITT) analysis, whereby all sample members are included in the analysis regardless of their participation in the program. This approach will yield an estimate of the program's average impact among 9th grade students given the opportunity to participate in the *Teen PEP* workshops, which is the most relevant estimate for a voluntary program. The magnitude of the ITT estimates could be affected by incomplete or non-participation by treatment group members and/or by participation of control group members in other pregnancy prevention services, making these factors important to consider in interpreting the

findings. To identify these factors, program staff will track attendance at each *Teen PEP* workshop for members of the treatment group, and data from all sample members on the receipt of other pregnancy-related services will be collected through the follow-up surveys and interviews conducted as part of the implementation evaluation. (See Chapter I for a more detailed description of our analytic approach.)

Data Source

All measures will be based on self-reported data drawn from a baseline survey, completed shortly before the program begins in the treatment schools, and two follow-up surveys, completed 6 and 18 months after program completion. The baseline and follow-up surveys will be paper-and-pencil questionnaires, administered in participating schools by field data collectors trained by Mathematica and Abt Associates Inc. For those students who miss the school-based survey administration, data will be collected through telephone interviews conducted by trained Mathematica interviewers. Monetary incentives will be provided for completion of the follow-up surveys. A $10 incentive payment will be provided for follow-up surveys completed during the school-based group survey administration and a $25 incentive payment will provided for surveys completed over the phone, due to the additional burden placed on individuals to complete a telephone survey outside of school hours.

Outcome Measures

Drawing on the survey data, the study team will construct, and estimate impacts for, a range of outcome measures (Table VII.2). These measures fall into two broad types: (1) sexual risk outcomes, which include both measures of sexual behavior and consequences of this behavior, most notably pregnancy; and (2) intermediate outcomes, which correspond to the mediating factors through which the program would most likely impact behavior (see Figure VII.1 above).

Two of the sexual risk outcomes shown in Table VII.2 will serve as "confirmatory measures of program impacts: (1) prevalence of sexual intercourse and (2) prevalence of unprotected sexual intercourse.[30] Delaying sexual initiation and increasing the use of birth control methods among sexually active teens are key goals of *Teen PEP*. Differences in the prevalence of sexual intercourse and unprotected sexual intercourse between members of the treatment and control groups will indicate whether the program was successful in achieving these goals. These primary outcome measures will be evaluated at both follow-ups, though impacts on behavior are more likely to emerge at the long-term, or 18-month, follow-up. While the ultimate goal of *Teen PEP* is to reduce teen pregnancy, any impacts of *Teen PEP* on this outcome are unlikely to emerge within the 18-month follow-up period and so it cannot serve as a confirmatory outcome for the study.

[30] To control for multiple hypothesis testing (the increased chance of falsely identifying an impact as statistically significant when examining effects on many outcomes), we are limiting the primary research question to two sexual risk outcomes, which are the focus of stated program objectives.

Table VII.2. Planned Outcomes for Measuring Program Impacts *(Teen PEP)*

Sexual Risk Outcomes	Intermediate Outcomes
Behavior: - *Prevalence of sexual intercourse* - *Prevalence of unprotected sex* - Prevalence of oral sex - Prevalence of anal sex - Frequency of sexual activity - Number of sexual partners **Consequences:** - Pregnancy - STDs	- Knowledge about sexual health, contraceptives, and HIV/STD prevention - Attitudes toward sexual intercourse and pregnancy - Communication about sexual activity, pregnancy, contraceptive use, and STDs - Perceptions of pregnancy and STD risk - Self-efficacy about sexual refusal and STD prevention and testing - Alcohol and drug use

Note: Outcomes shown in italics reflect the confirmatory outcomes for the impact analysis.

Beyond this narrow, confirmatory impact analysis, a comprehensive exploratory analysis will estimate and assess impacts across the range of other outcomes shown in Table VII.2, including both the remaining sexual risk outcomes (shown in the left column) and the full set of intermediate outcomes (shown in the right column). Although these outcomes will not be the basis for confirmatory evidence on the effects of the program, they will provide a complete picture of potential effects of the program on teen sexual risk behaviors and their consequences, and pathways through which the program may affect sexual behaviors. In particular, findings from the analysis of intermediate outcomes, or mediating factors, will help us understand the sources and nature of any confirmatory evidence of program impacts on the prevalence on sexual intercourse and unprotected sex. Exploratory analyses will also be conducted to understand whether and how program impacts vary by key subgroups, such as gender and/or race/ethnicity.

Sample Size and Statistical Power

As a means of summarizing the statistical power afforded by our sample design, Table VII.3 shows the estimated minimum detectable impacts (MDIs) for the two confirmatory outcomes under different assumptions. As the unit of random assignment is schools, the study's ability to detect statistically significant impacts will depend on the level of correlation in the outcomes among students within the same schools (intraclass correlation coefficient [ICC]). To assess the sensitivity of the MDIs to the level of correlation, we provide estimates assuming a relatively low (0.01) and high (0.035) ICC. In addition, for each level of the ICC, we provide an estimate of the MDI based on more and less optimistic assumptions about the explanatory power (R^2) of the regression models used to estimate the impacts.

Naturally, as seen in the table, the lowest MDI for each outcome arises with assumptions of a low ICC and high R^2, whereas the highest MDI arises given the opposite assumptions. Specifically, for the prevalence of sexual intercourse (projected to have a control group mean of 55 percent), the study is powered to detect a minimum of 7.3 percentage points to 13.3 points depending on the assumed ICC and R^2.[31] For prevalence of unprotected sex (with an anticipated control mean of 10

[31] Prevalence of sexual intercourse is defined as the proportion of youth who ever had sexual intercourse. An approximate mean for the control group is estimated based on data from the New Jersey and North Carolina 2009 Youth Risk Behavior Survey , and represents the average of the proportion that reported ever having sexual intercourse between the two states, among teens in grade 11.

percent), the MDI ranges from 4.4 to 8.0 percentage points. Both of these ranges reflect relatively substantial impacts, though within levels that might be expected of the program given the scale of its investment in the schools it serves and the existing evidence base on other teen pregnancy prevention programs (Clark et al 2005, Coyle et al. 2004).

Table VII.3. Minimum Detectable Impacts for Illustrative Outcomes *(Teen PEP)*

				Minimum Detectable Impacts	
				R^2 = High Estimate[a]	R^2 = Low Estimate[b]
	ICC	Sample Size (Cluster/Ind.)	Mean Outcome (%)	PPΔ	PPΔ
Second Follow- Up:					
Prevalence of sexual intercourse	Low[c]	16/2,778	55	7.3	8.3
	High[d]	16/2,778	55	11.7	13.3
Prevalence of unprotected sex	Low	16/2,778	10	4.4	5.0
	High	16/2,778	10	7.1	8.0

PPΔ=percentage point change. Minimum detectable impacts are estimated for a 5 percent significance level using a two-tailed test and 80 percent power. The equation used to calculate the minimum detectable impacts for cluster-level random assignment is:

$$MDI = 2.8 \times \sigma \sqrt{\rho \times \left(1 - R_c^2\right) \times \left(\frac{1}{C_T} + \frac{1}{C_C}\right) + (1 - p) \times \left(1 - R_i^2\right) \times \left(\frac{1}{N_T} + \frac{1}{N_C}\right)},$$

where σ is the standard deviation of the outcome measure, ρ is the estimated ICC, C_T and C_C are the number of clusters in the treatment and control groups, and N_T and N_C are the number of individuals in the treatment and control groups. The estimated sample size is based on an 85 percent retention rate for the second follow-up.

[a]High estimate of R-squared is 0.3.

[b]Low estimate of R-squared is 0.1.

[c]Low estimate of ICC is 0.01.

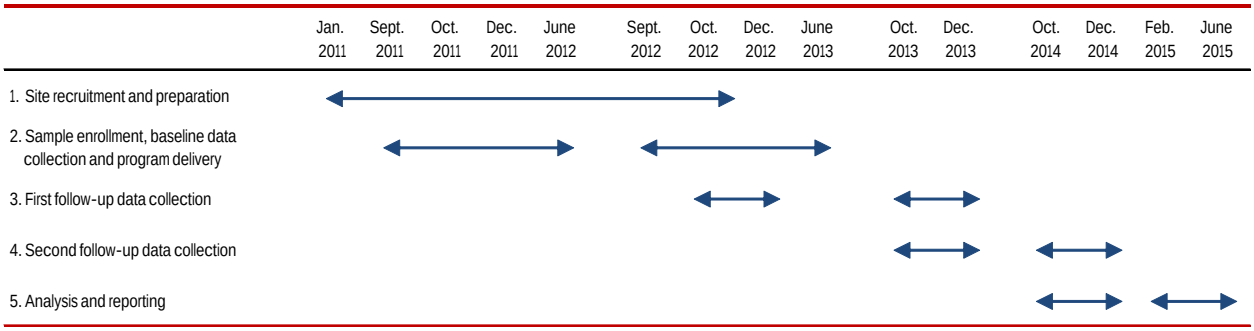
[d]High estimate of ICC is 0.035.

Evaluation Timeline and Progress to Date

The timeline for the *Teen PEP* evaluation extends from start-up planning in 2011 through final results in 2015 (Figure VII.2). Recruitment of the first study schools began in early 2011 and continues in both states. Sample enrollment, baseline data collection, and program delivery are under way in 2011–2012 for the four North Carolina schools initially recruited and are expected to take place for the remaining study schools in both states in 2012–2013. (As noted previously, recruitment of these schools is ongoing.)

The first and second follow-up data collection will occur in October to December of 2012 and 2013, respectively, for the four North Carolina schools recruited as part of the first cohort. In the remaining 12 schools (8 of which have already been recruited), follow-up data collection is expected to occur in October to December of 2013 and 2014, respectively. A final report presenting the results of the *Teen PEP* evaluation is expected at the end of the second quarter of 2015, though this date may be revised to the second quarter of 2016 should recruitment of the remaining four schools for the study need to extend into a third year. An interim impact analysis will be conducted based on the first follow-up data and reported in an interim report in late 2014.

Figure VII.2. Timeline for the Evaluation of *TEEN PEP*[a]



[a]Timeline assumes recruitment of the targeted 16 study schools will be completed in two cohorts, as shown.

VIII. EVALUATION OF *WAIT TRAINING*

Rigorous evaluations of programs promoting abstinence until marriage are relatively scarce, offering little evidence on their effectiveness in reducing sexual risk behavior among adolescents. The Teen Pregnancy Prevention Research Evidence Review (PPRER), funded by HHS, identified only six evaluations of abstinence until marriage programs published between 1989 and 2010 that used a random assignment method. The total number of experimental evaluations of teen pregnancy prevention programs published during the same period is approximately 130.[32] As many school districts in the United States choose abstinence until marriage curricula for their sex education programs, there is a need to improve the evidence base for this approach. In an effort to fill this gap, Mathematica, in collaboration with Live the Life Ministries (LtL) and Dr. Frank Fincham of Communicate LLC, is conducting an experimental evaluation of *WAIT Training*, an abstinence-until-marriage program designed for middle school students that aims to provide adolescents with the knowledge, skills, and confidence to delay sexual initiation, thereby reducing the prevalence of teen pregnancy and STDs.

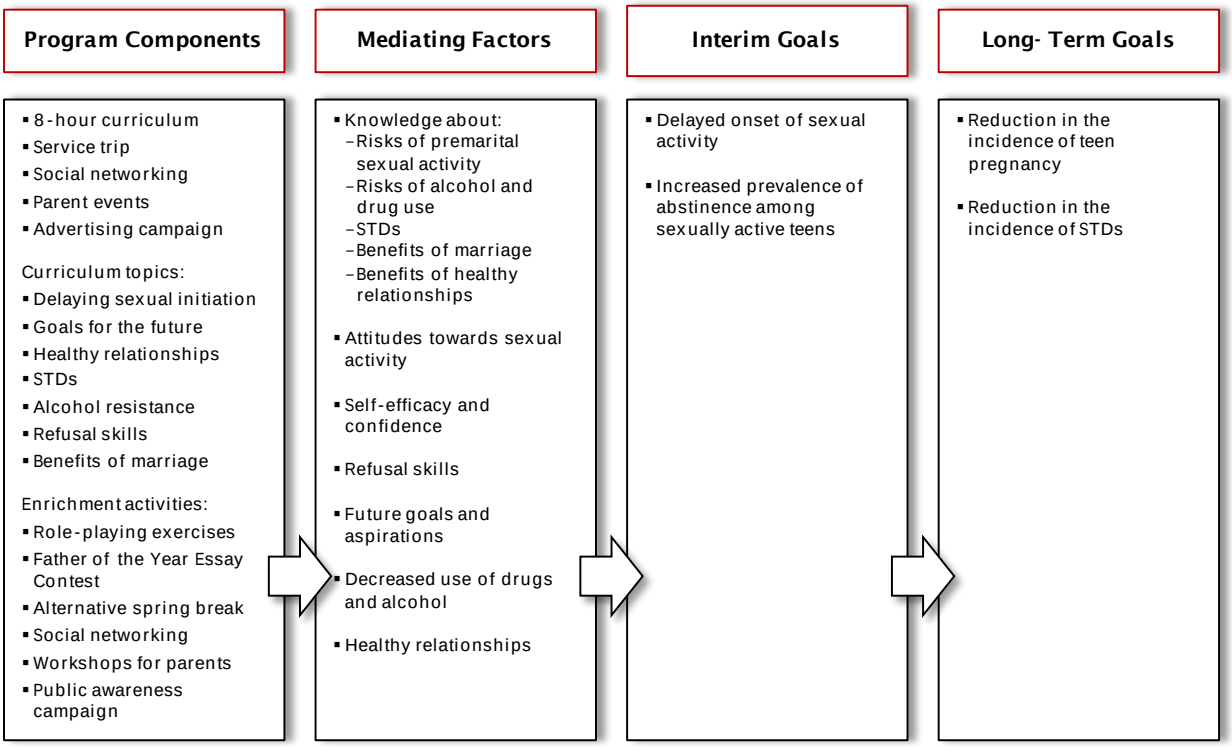
A. Program Features and Evaluation Setting

The logic model of the *WAIT Training* program (Figure VIII.1) is grounded in the Theory of Planned Behavior, which states that individuals' behaviors are influenced by their perceived control over their actions as well as their attitudes and beliefs (Ajzen 1991). Drawing on elements of this theory, *WAIT Training* aims to educate middle school and high school students on the risks of sexual activity outside of marriage, and to shape students' attitudes, beliefs, and ultimately their behavior. The main component of *WAIT Training* is a classroom-based curriculum designed for youth in grades 6 to 12. The program also offers activities outside of the classroom that are designed to reinforce the messages of the curriculum, to stay connected with participating students, and to get parents, teachers, and other students involved.

The curriculum to be tested in this evaluation will consist of eight lessons, selected from a larger compendium of *WAIT Training* curriculum materials that include seven core units and over 100 exercises. The selected lessons will be implemented as eight 1-hour sessions delivered over eight consecutive school days per year. Although the typical delivery of *WAIT Training* allows schools to tailor the intervention by selecting lessons and the grades in which they will be delivered, for this evaluation schools will be required to follow the same lesson plan. Classroom instruction is divided by sex, with one trained teacher delivering the program for girls and one for boys. The curriculum is usually delivered in health, physical education, or science classes.

[32] Authors' analysis based on findings from HHS' Teen Pregnancy Prevention Research Evidence Review (PPRER).

Figure VIII.1. Logic Model of the *WAIT Training* Intervention



The seven units of the *WAIT Training* curriculum are the following:

1. ***Learning About Yourself and Others.*** The objective of the unit is to help students learn about themselves and their peers; discover their own skills, talents, hopes, and dreams; and understand and respect individualism as well as diversity.
2. ***Friendship, Dating and Love.*** The unit aims to teach students the characteristics of a healthy relationship, and how to develop healthy, fulfilling, satisfying, and committed relationships.
3. ***Cultural Influences.*** This unit is designed to educate students about various cultural influences; examples of topics covered are: the influence of media on their lives, alcohol consumption, and vulnerability to sexual advances.
4. ***Differences Between Men and Women.*** The unit aims to educate students about the physical, emotional, and behavioral differences between men and women with the goals of improving their understanding and appreciation of the differences and improving communication between the genders. The ultimate goal is to help them pursue healthier relationships.
5. ***Consequences of Teen Sex.*** The objective is to help students understand the physical, emotional, and financial ramifications of premarital sexual activity, including information about the transmission, prevention, and treatment of STDs.
6. ***Dealing Effectively with Conflict.*** The unit includes lessons and exercises that aim to help students learn and practice sexual refusal skills and conflict resolution.

7. ***Commitment and Marriage***. The purpose of the unit is to increase students' value of the institution of marriage. Emphasis is placed on both the benefits of being and staying married and the many negative health and social outcomes associated with fractured or never-married families.

The eight lessons selected by LtL for the intervention cover topics from all seven units. The lessons are anchored in the program's logic model and, together, they aim to demonstrate the consequences of premarital sex and the benefits of waiting until marriage, to educate students about the physical risks of sexual activity (such as STDs), and to encourage students to engage in healthy relationships. Each lesson involves interactive activities such as role-playing, completion of worksheets, and various exercises.

In addition to classroom instruction, *WAIT Training* can include various activities that take place outside of the classroom, although these will not be part of the intervention tested here. The purpose of the extra-curricular activities is to maintain contact with participating students and to involve more students and the larger community in the program. For example, a Father of the Year essay contest and quarterly seminars for parents seek to build parent and community involvement.

Evaluation Sample and Setting

The evaluation will be conducted in approximately 17 middle schools and 16 high schools located in central and northern Florida and southern Georgia. These schools will be grouped into 16 "clusters," each of which will include a high school and its feeder middle school(s). A first cohort, consisting of 10 clusters, has already been recruited with programming scheduled to begin in the spring of the 2011–2012 school year. A second cohort, consisting of six clusters, is currently being recruited with programming scheduled to begin in the spring of the 2012–2013 school year. Students enrolled in 7th grade during the 2011–2012 (for the first cohort) and the 2012–2013 academic year (for the second cohort) and who consent to participate in the evaluation will comprise the study sample.

Following the experimental design for the study, half of the school clusters will be randomly assigned to a "treatment group" that will provide *WAIT Training* to students in the 7th and 8th grades, while the other half will be randomly assigned to a "control group" that will provide a program focused on diet, exercise, and nutrition called L.E.A.N. (Lifestyle, Exercise, Attitude, Nutrition), also in the 7th and 8th grades. In most schools, these programs will be delivered as part of a required health, science, or physical education class.

All schools participating in the evaluation are expected to be located in counties with a relatively disadvantaged population and a high teen birth rate. The 10 clusters recruited to date are located in six counties, each with a predominantly rural, lower-income population composed of mostly white non-Hispanics and African Americans. In 2010, an estimated 11.6 percent of families living in these counties had incomes below the federal poverty level.[33] In the 2010–2011 school year, about 49

[33] Based on author's calculations from statistics in the U.S. Census Bureau 2006-2010 American Community Survey.

percent of students in participating schools received free or reduced-price lunches.[34] The average teen birth rate in the six counties in 2010 was 52 per 1,000 women, compared to 46 per 1,000 women for the entire state of Florida.[35]

B. Evaluation Design

Based on the experimental design, the main difference between youth in the treatment and control groups is that those in the treatment group will receive the *WAIT Training* curriculum while those in the control group will not (Table VIII.1). Youth in the control group will receive an alternative curriculum, project L.E.A.N., but its focus is on diet and exercise and so is not expected to affect the outcomes of main interest to this study. Thus, the difference in the outcomes of interest—most notably, reported sexual abstinence—between the treatment and control groups will provide a valid estimate of *WAIT Training*'s impact among study youth. Sample members in both study groups can access existing services offered to teens during the study period, some of which may address risk factors related to teen pregnancy (Table VIII.1). Available information suggests that few, if any, school-based teen pregnancy prevention services are available to study youth (beyond access to *WAIT Training* among youth in the treatment group). This may be due to the fact that Florida does not mandate sex education in schools; the state only imposes requirements on the content of sex education programs if schools choose to offer one.[36]

Table VIII.1. Summary of Contrast Between the Treatment and Control Conditions *(WAIT Training)*

	Treatment Group	Control Group
Intervention	Eight core lessons of the *WAIT Training* curriculum	L.E.A.N. curriculum
Existing Services	Other services and programs available in the selected Florida counties related to teen pregnancy prevention (no specific services offered by participating schools have been identified)	

Because the impact of *WAIT Training* may be influenced by similar programming and services that may be available to participating students, it is critical that we measure the availability of "existing services" in both schools and the broader community, so we can accurately characterize the actual treatment-control contrast and interpret the study's findings. Throughout the study, the evaluation team will remain in contact with school staff to confirm that competing services are not introduced into the study schools, a step that could undermine the study's value. In addition, information about service receipt will come from two main sources: (1) questions on the participant follow-up surveys conducted as part of the impact study and (2) three focus groups conducted as

[34] Based on author's calculations from statistics in Common Core of Data (CCD) Public School Data for the 2009-2010 and 2010-2011 school years, available at: http://nces.ed.gov/ccd/bat/.

[35] Based on author's calculations from statistics in the U.S. Census Bureau 2010 Census Summary File 1 and the county health rankings. http://www.countyhealthrankings.org. Accessed February 6, 2012.

[36] Florida requires that the content of sex education classes stress the negative outcomes of teen sexual activity and the importance of delaying sexual activity until marriage, that sex education be age appropriate, and that parents have an option to opt out of the program.

part of an in-depth site visit to understand the program and its context—two with members of the treatment group and one with members of the control group. Interviews with program stakeholders and other key informants, also conducted as part of the site visit, will also help us to describe the backgrounds and experiences of participating students and the kinds of services and activities that they may access in the community.

Sample Intake and Random Assignment

At each middle school recruited to participate in the evaluation, the study team will collect parental consent from all 7th grade students who would be eligible for *WAIT Training* were it to be offered in their school. Consent will be gathered with the assistance of school staff, who will provide the evaluation team with copies of student rosters and will help distribute consent forms to students and collect forms after parents have signed them.

To ensure the integrity of the study design, the random assignment of schools clusters will take place after the completion of consent gathering in spring 2012 for the first cohort of 10 schools and in fall 2012 for the second cohort of 6 schools. To increase precision of the impact estimates, we will employ a stratified random assignment design, using cluster-level characteristics such as the racial composition of the student population and study consent rate as bases for grouping school clusters.

Measuring Program Impacts

Impacts of the *WAIT Training* curriculum will be evaluated in the 9th and 10th grades—the first two grades after the program has been completed in the treatment group middle schools. All estimates will be based on an intent-to-treat (ITT) analysis, meaning it will include all youth in the study sample, including those in the treatment group who may ultimately not participate in the program We will use this approach because it provides an estimate of the average program impact among all program-eligible students (who are participating in the study), which is the most relevant estimate for a mandatory in-school program.

Given the random assignment design, valid impacts can be estimated by simply comparing the post-intervention outcomes between the treatment and control groups. However, in order to increase precision, we will adjust the impact estimates for any baseline differences in students' characteristics and reported behaviors using regression models. To account for the possible correlation in the outcomes among students within the same cluster (intraclass correlation), we will use the Huber-White method to estimate the standard error associated with each impact estimate.

Data

All measures will be based on self-reported data collected at three points: (1) prior to random assignment, approximately 2 months before program delivery; (2) at first follow-up, approximately 20 months after baseline and 4 months after completion of the second year of the intervention; and (3) at second follow-up, approximately 32 months after baseline or 16 months after completion of the second year of the intervention. Trained Mathematica field staff will administer the surveys during the school day, and the majority of students will complete a paper-and-pencil questionnaire in a group setting. Students who are not able to complete follow-up surveys during the school day will be asked to complete a telephone-based survey to minimize sample attrition.

Outcome Measures

The study team will construct and estimate impacts for a range of outcome measures (Table VIII.2). These measures fall into two broad types: (1) sexual risk outcomes, which include measures of both abstinence and sexual risk behavior; and (2) intermediate outcomes, which correspond to the mediating factors through which the program would most likely impact behavior (see Figure VIII.1 above).

The two confirmatory outcomes for the evaluation are: (1) prevalence of abstinence from sexual intercourse, and (2) prevalence of unprotected sex[37]. Given its exclusive focus on abstinence, *WAIT Training* does not aim to affect rates of unprotected sex. However, it could affect this outcome indirectly, making it fundamental to assess in considering the effects of the program. At each follow-up, the prevalence of abstinence will be measured as the proportion of students who did not have sexual intercourse. Prevalence of unprotected sex will be measured as the proportion of students who had sex without using an effective method of contraception during the months prior to each follow-up.

In addition to the confirmatory impact analysis, a more extensive analysis will assess impacts across a range of sexual risk behaviors and behavioral consequences (shown in the left column of Table VIII.2). Although these outcomes will not be the basis for confirmatory evidence on the effects of the program, they will provide a complete picture of potential effects of the program on teen sexual risk behaviors and the consequences. In addition, the exploratory analysis will estimate impacts on a range of mediating factors (shown in the right column of Table VIII.2). Examples of these mediators include youths' views on sexual intercourse prior to marriage, their intentions to engage in sexual activity in the near future or before marriage, their knowledge of STDs and pregnancy risk, and their refusal skills.

Table VIII.2. Planned Outcomes for Measuring Program Impacts *(WAIT Training)*

Sexual Risk Outcomes	Intermediate Outcomes
Behavior:	- Views about sexual intercourse at current age
- *Prevalence of abstinence from sexual intercourse*	- Views about sexual intercourse before marriage
- *Prevalence of unprotected sex*	- Intention to have sexual intercourse before marriage
- Prevalence of oral sex	- Intention to use condoms if sexually active
- Frequency of sexual activity	- Refusal skills
- Number of sexual partners	- Knowledge of the risk of STDs and pregnancy
Consequences:	
- Pregnancy	
- STDs	

Note: Outcomes shown in italics reflect the confirmatory outcomes for the impact analysis.

[37] While the program's most vital long-term goal is to reduce the incidence of teen pregnancy, it is not included as a confirmatory outcome because the study is not expected to have sufficient statistical power to detect an impact on it. The main reason is that the incidence of pregnancy is expected to remain low in the control group by the time of the final follow-up survey (in 10th grade), making it unlikely that a detectable impact on this outcome could emerge during the study period.

Sample Size and Statistical Power

As the unit of assignment for the evaluation is a cluster of schools, the study's ability to detect statistically significant impacts will be influenced by the level of correlation in the outcomes among students within the same cluster. In Table VIII.3, we provide minimum detectable impacts (MDIs) for the two confirmatory outcomes: (1) prevalence of sexual abstinence, and (2) prevalence of unprotected sex, both measured at second follow-up. To assess the sensitivity of the MDIs to the level of correlation among students in the same clusters, we provide estimates based on low (0.01) and high (0.035) intraclass correlation coefficients (ICCs) and based on high and low levels of the regression R-square.

The MDIs shown in Table VIII.3 indicate that the study should be sufficiently powered, though less so for the unprotected sex outcome, particularly if assumptions about the R-squared and ICC prove optimistic. For the first outcome, abstinence from sex, the estimated MDI ranges from about 8 to 13 percentage points (depending on the ICC and R-squared assumptions) given an anticipated control group mean of 50 percent.[38] For the second outcome, prevalence of unprotected sex, the estimated MDI ranges from 5 to nearly 9 percentage points given an anticipated control group mean of 10 percent. While these ranges reflect sizeable effects for both outcomes, the one for abstinence is at or below the impacts found for some school-based programs in prior studies (Clark et al. 2005, Coyle et al. 2006). This suggests that, even if the study experiences a relatively high ICC, it will be suitably powered for detecting impacts on the program's main goal—increasing the rate of sexual abstinence among teens. This may not be the case for the unprotected sex outcome, however, where the MDI under a high ICC reflects nearly a 90 percent (8.7 point) reduction from the estimated control group mean of 10 percentage points.

Evaluation Timeline and Progress to Date

The evaluation will span a five-year period, from spring 2011 through the end of 2015 (Figure VIII.2). During the first year, the study team finalized plans for the evaluation design, developed the survey instruments, recruited a first cohort of 12 school clusters, enrolled) eligible students from participating schools, and administered the baseline survey to them. Recruitment of a second cohort of four more school clusters is underway and will continue through fall 2012. Following school recruitment, consent gathering for students in the second cohort will occur in early spring 2013, followed by baseline data collection. Random assignment of the clusters in the first cohort occurred in spring 2012; for the second cohort random assignment is planned for spring 2013, after completion of baseline data collection. The first and second follow-up surveys will be administered in fall 2013 and fall 2014, respectively, for the first cohort of students, and in fall 2014 and fall 2015 for the second cohort.

[38] For both outcome measures, an approximate mean for the control group was estimated based on data from the sample of 10th graders in the 2009 Florida Youth Risk Behavior Survey (YRBS). The mean for the prevalence of sexual intercourse represents the proportion of youth who reported ever having sexual intercourse. The mean for the prevalence of unprotected sex represents the proportion of the full sample who had sexual intercourse in the last three months and did not use a condom during last sexual encounter.

Table VIII.3. Minimum Detectable Impacts for Illustrative Outcomes *(WAIT Training)*

				Minimum Detectable Impacts	
				R^2 = High Estimate[a]	R^2 = Low Estimate[b]
	ICC	Sample Size Cluster/Ind.)	Mean Outcome (%)	PPΔ	PPΔ
Second Follow- Up:					
Prevalence of abstinence from sexual intercourse	Low[c]	16/1700	50	8.1	9.2
	High[d]	16/1700	50	12.3	13.9
Prevalence of unprotected sex	Low	16/1700	10	5.1	5.8
	High	16/1700	10	7.7	8.7

PPΔ=percentage point change. Minimum detectable impacts are estimated for a 5 percent significance level using a two-tailed test and 80 percent power. The equation used to calculate the minimum detectable impacts for cluster-level random assignment is:

$$MDI = 2.8 \times \sigma \sqrt{\rho \times \left(1 - R_c^2\right) \times \left(\frac{1}{C_T} + \frac{1}{C_C}\right) + (1 - p) \times \left(1 - R_i^2\right) \times \left(\frac{1}{N_T} + \frac{1}{N_C}\right)},$$

where σ is the standard deviation of the outcome measure, ρ is the estimated ICC, C_T and C_C are the number of clusters in the treatment and control groups, and N_T and N_C are the number of individuals in the treatment and control groups. The estimated sample size is based on an 85 percent retention rate for the second follow-up.

[a]High estimate of R-squared is 0.3.

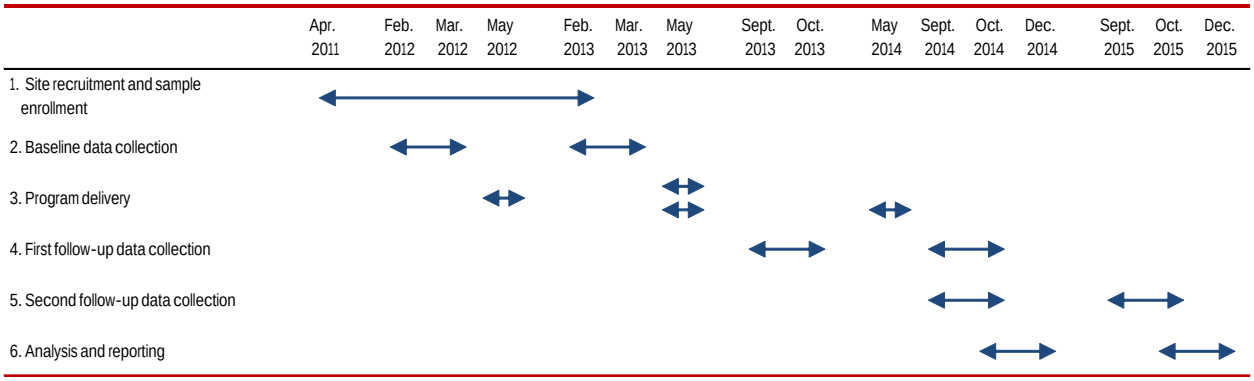
[b]Low estimate of R-squared is 0.1.

[c]Low estimate of ICC is 0.01.

[d]High estimate of ICC is 0.035.

A final report, summarizing impacts of *WAIT Training* across the range of confirmatory and exploratory outcomes, will be completed in December 2015. Data for this report will be obtained from all three student surveys, the baseline survey in 7th grade and the first and second follow-up surveys in the 9th and 10th grades. In addition to this report, an interim report will likely be completed in late 2014, summarizing preliminary impact findings based on only the first two rounds of data collection (baseline and first follow-up).

Figure VIII.2. Timeline for the Evaluation of *WAIT Training*



REFERENCES

Ajzen, I. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes*, vol.50, 1991, pp. 179–211.

Allen, M., S. Miller, and J. Abbey, J. "Teens in Foster Care: Preventing Pregnancy and Building Self-Sufficiency." Washington, DC: Adolescent Pregnancy Prevention Clearinghouse, Children's Defense Fund, 1987.

Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society,* Series B (Methodological), vol. 57, no. 1,1995, pp. 289–300.

Bilaver, L.A., and M.E. Courtney. "Foster Care Youth." Science Says #27. Washington, DC: The National Campaign to Prevent Teen Pregnancy, 2006.

Bloom, H. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review,* vol. 8, no. 2, 1984, pp. 225–246.

Boland, J., B. Lian, and C. Formichella. "The Origins of Hopelessness Among Inner-City African-American Adolescents." American Journal of Community Psychology, vol. 36, no. 3/4, 2005, pp. 293–305.

Centers for Disease Control and Prevention. "2009 Sexually Transmitted Disease Surveillance:" Atlanta, GA: CDC, 2010.

Centers for Disease Control and Prevention. "Youth Risk Behavior Surveillance—United States, 2009." *MMWR,* vol. 59, no. SS-5, 2010a, pp. 1–142.

Centers for Disease Control and Prevention. "2009 Sexually Transmitted Diseases Surveillance." Atlanta, GA: CDC, 2010b. Available at http://www.cdc.gov/std/stats09/toc.htm. Accessed January 5, 2012.

Clark, L.F., K.S. Miller, S.S. Nagy, J. Avery, D.L. Roth, N. Liddon, and S. Mukherjee. "Adult Identity Mentoring: Reducing Sexual Risk for African-American Seventh Grade Students." Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine, vol. 37, no. 4, 2005, pp. 337e1–337e10.

Conde-Agudelo A., A. Rosas-Bermudez, and A.C. Kafury-Goeta. "Birth Spacing and Risk of Adverse Perinatal Outcomes: A Meta Analysis." Journal of the American Medical Association, vol. 295, 2006, pp. 1809–1823.

Connell, R. Gender and Power: Society, the Person and Sexual Politics. Palo Alto, CA: University of California Press, 1987.

"County Health Rankings." 2011. Available at: http://www.countyhealthrankings.org. Accessed December 14, 2011.

Courtenay, W.H. "Constructions of Masculinity and Their Influence on Men's Well-being: A Theory of Gender and Health." Social Science and Medicine, vol. 50, no. 10, 2000, pp. 1385–1401.

Coyle, K. K., D. B. Kirby, B. V. Marin, C. A. Gomez, and S. E. Gregorich. "Draw the Line/Respect the Line: A Randomized Trial of a Middle School Intervention to Reduce Sexual Risk Behaviors." American Journal of Public Health, vol. 94, 2004, pp. 843–851.

Coyle, K.K., D.B. Kirby, L.E. Robin, S.W. Banspach, E. Baumler, and J.R. Glassman. "All4You! A Randomized Trial of an HIV, Other STDs, and Pregnancy Prevention Intervention for Alternative School Students." AIDS Education and Prevention, vol. 18, no. 3, 2006, pp. 187–203.

DiClemente, R.J., G.M. Wingood, E.S. Rose, J.M. Sales, D.L. Lang, A.M. Caliendo, J.W. Hardin, and R.A. Crosby. "Efficacy of Sexually Transmitted Disease/Human Immunodeficiency Virus Sexual Risk-Reduction Intervention for African American Adolescent Females Seeking Sexual Health Services." Archives of Pediatric & Adolescent Medicine, vol. 163, no. 12, 2009, pp. 1112–1121.

Flay, B. R., S. Graumlich, E. Segawa, J. L. Burns, and M. Y. Holliday. "Effects of 2 Prevention Programs on High-Risk Behaviors Among African American Youth: A Randomized Trial." Archives of Pediatrics & Adolescent Medicine, vol. 158, no. 4, 2004, pp. 377–384.

Furstenberg, F.F. Jr., J. Brooks-Gunn, and S.P. Morgan. Adolescent Mothers in Later Life. New York: Cambridge University Press, 1987.

Glanz, K., B.K. Rimer, and F.M. Lewis. Health Behavior and Health Education: Theory, Research and Practice. San Francisco: Wiley & Sons, 2002.

Goesling, Brian, and Chris Trenholm. "Pregnancy Prevention Research Evidence Review (PPRER): Lessons Learned and Recommendations." Princeton, NJ: Mathematica Policy Research, 2010.

Hampton, T. "Researchers Seek Ways to Stem STDs: 'Alarming' STD Rates Found in Teenaged Girls." *JAMA,* vol. 299, no. 16, 2008, pp. 1888–1889.

Hoffman, S.D. *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy.* Washington, DC: The Urban Institute Press, 2008.

Ickovics, J.R., and R. Rodin. "Women and AIDS in the United States; Epidemiology, National History and Meditation Mechanisms." Health Psychology, vol. 11, 1992, pp. 1–16.

Ikramullah, E., Keith, J., Manlove, J., and K. Anderson Moore. "Table 3. Number of Births to Teenage Mothers in Large Cities by Age and by Race/Ethnicity, Number and Percentage of Births to Unmarried Mothers under Age 20, Percentage of Births to Mothers 15-19 that are Repeat Births (Second or Higher Order Births), and Percentage of all Births in City to Mothers under Age 20, 1991 & 2005." Facts at a Glance, Publication no. 29, 2008, p. 4.

Institute for Education Sciences, National Center for Education Statistics, U.S. Department of Education. "Common Core of Data (CCD) Public School Data for the 2009-2010 and 2010-2011 School Years." Available at: http://nces.ed.gov/ccd/bat/

Jemmott, J.B., L.S. Jemmott, P.K. Braverman, and G.T. Fong. "HIV/STD Risk Reduction Interventions for African American and Latino Adolescent Girls at an Adolescent Medicine Clinic: A Randomized Controlled Trial." Archives of Pediatrics & Adolescent Medicine, vol. 159, no. 5, 2005, pp. 440–449.

Kalmuss, D.S., Namerow, P.B. "Subsequent Childbearing Among Teenage Mothers: The Determinants of a Closely Spaced Second Birth." Family Planning Perspectives, vol. 26, no. 4, 1994, pp. 149-159.

Kandrack, M.A., K.R. Grant, and A. Segall. "Gender Differences in Health Related Behavior: Some Unanswered Questions." Social Science and Medicine, vol. 32, no. 5, 1991, pp. 579–590.

Klerman, L.V. "Another Chance: Preventing Additional Births to Teen Mothers." Washington, DC: National Campaign to Prevent Teen Pregnancy, 2004.

Kost, K., S. Henshaw, and L. Carlin. "U.S. Teenage Pregnancies, Births and Abortions: National and State Trends and Trends by Race and Ethnicity." New York: Guttmacher Institute, 2010. Available at http://www.guttmacher.org/pubs/USTPtrends.pdf

Kost, K., S. Henshaw, and L. Carlin. "U.S. Teenage Pregnancies, Births and Abortions: National and State Trends and Trends by Race and Ethnicity." New York: Guttmacher Institute, 2010.

Kramer, C. Y. "Extension of the Multiple Range Test to Group Means with Unequal Numbers of Replications." *Biometrics,* vol. 12, 1956, pp. 307–310.

Los Angeles County Department of Public Health, Office of Health Assessment and Epidemiology. "Key Indicators of Health, by Service Planning Area," June 2009.

Martin, J.A., B.E. Hamilton, S.J. Ventura, M.J.K. Osterman, S. Kirmeyer, T.J. Mathews, and E.C. Wilson. "Births: Final Data for 2009." National Vital Statistics Reports, vol. 60, no 1. Hyattsville, MD: National Center for Health Statistics, 2011.

Martin, Joyce, Brady E. Hamilton, Stephanie J. Ventura, and Michelle J.K. Osterman. "Births: Final Data for 2009." *National Vital Statistics Reports*, vol. 60, no. 1, November 2011, pp. 1–104.

Matts, John P. and John M. Lachin. "Properties of Permuted-Block Randomization in Clinical Trials." *Controlled Clinical Trials*, vol. 9, 1988, pp. 327–344.

National Center for Health Statistics, Centers for Disease Control and Prevention. "Natality 2009." Downloadable public use file; annual internet product. Hyattsville, MD: NCHS, 2011. Available at: http://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm

Nationwide Children's Hospital. "Full Potential Community Report: Teen Pregnancy Trends." Columbus, OH: Nationwide Children's Hospital. Available at: http://www.nationwidechildrens.org/full-potential-teen-pregnancy

Oklahoma Institute for Child Advocacy and the University of Oklahoma National Resource Center for Youth Services. "POWER Through Choices: Sexuality Education for Youth in Foster and Group Care." OICA and NRCYS, 2010.

Pleck, J.H., F.L. Sonenstein, and L.C. Ku. "Masculinity Ideology: Its Impact on Adolescent Males' Heterosexual Relationships." Journal of Social Issues, vol. 43, no. 3, 1993, pp. 11–29.

Schochet, Peter Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review*, vol. 33, no. 6, December 2009.

Schulz, Kenneth F. and David A. Grimes. "Generation of Allocation Sequences in Randomised Trials: Chance, Not Choice." *Lancet,* vol. 356, 2002, pp. 515–519.

Stewart, J. "The Mommy Track: The Consequences of Gender Ideology and Aspirations on Age at First Motherhood." Journal of Sociology and Social Welfare, vol. 30, no. 2, 2003, pp. 3–30.

Tukey, J. W. "The Problem of Multiple Comparisons." In *Mimeographed notes*. Princeton, NJ: Princeton University, 1953.

United Nations. *2008 Demographic Yearbook*. New York: United Nations, 2010.

APPENDIX A

PPA PROGRAM IMPLEMENTATION AND DATA COLLECTION SCHEDULE

THIS PAGE IS INTENTIONALLY BLANK

Figure A.1. PPA Program Implementation and Data Collection Schedule

^Sample with rolling enrollment of sample members.

*Sites with two or more enrollment cohorts.

[a]Timeline assumes recruitment of the targeted 16 study schools will be completed in two cohorts, as shown.

THIS PAGE IS INTENTIONALLY BLANK

THIS PAGE IS INTENTIONALLY BLANK

MATHEMATICA
Policy Research

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research