

ATTACHMENT B
ANALYSIS PLAN FOR PREP IIS STUDY

ANALYSIS PLAN FOR PREP IMPACT AND IN-DEPTH IMPLEMENTATION STUDY

The purpose of the PREP Impact and In-Depth Implementation Study (IIS) is to rigorously assess the impacts and implementation of funded programs in four or five selected PREP sites. In each site, ACF expects to recruit and enroll a sample of 1,200 to 1,500 youth (for a total of 6,000 youth across four or five sites). Youth will be randomly assigned to a treatment group that receives the program being tested or to a control group that does not. Program impacts will be analyzed with survey data collected at baseline, 8-12 months post-program completion, and approximately 12 months after that. Program implementation will be analyzed based on data from site visits; interviews with program developers, program leaders and staff, and program partners and other stakeholders; an online survey of frontline staff and supervisors; and group interviews with participating youth.

Our analysis plan for the IIS has five main components: (1) a random assignment evaluation of each site using either an individual random assignment or cluster-based random assignment approach, (2) an early analysis of baseline data, (3) a primary impact analysis of key behavioral outcome measures, (4) exploratory impact analyses of secondary research questions, and (5) an in-depth implementation analysis. These are described below.

Random assignment. The impact study is expected to include four to five different sites. We expect that each site will focus on a different program model, population, and setting, reflecting the diversity of programs funded under PREP. Each site will be analyzed separately, so we are not constrained to using the exact same research design in each site. Rather, to achieve the most rigorous evaluation of each site, we will customize our approach to random assignment to the unique circumstances of each site. In some sites, we anticipate that random assignment of individuals will offer the best approach. In others, random assignment of schools or other types of clusters may be the best approach. Regardless of the specific approach, random assignment will be conducted only within each site. Because each site will be analyzed separately, we do not plan to conduct random assignment or make comparisons across sites.

Baseline analysis. As soon as baseline data collection has been completed in each site, we will begin preliminary analyses of the baseline data. We will use these analyses to describe the study sample in each site and compare it with the target population. We will also assess whether random assignment successfully generated treatment and control groups balanced on important baseline characteristics. To support this analysis, our baseline survey will collect key measures of demographics (such as age, gender, race, and ethnicity) and other personal characteristics (such as prior sexual experience) needed to describe the study sample and examine the equivalence of the treatment and control groups.

Primary impact analysis. Impact analysis will begin after the completion of follow-up data collection in each site. With a random assignment design, unbiased impact estimates can be obtained from the difference in unadjusted mean outcomes at follow up 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 treatment 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

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

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. These baseline characteristics will be gathered on baseline surveys. The parameter estimate for λ is the estimated impact of the program.

If clusters, 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

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

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 sexual partners), whereas logistic regression procedures will be specified for binary

outcomes (such as ever had sexual intercourse). Weights will be created for each site to account for any differences in random assignment, sampling, consent, or nonresponse probabilities among study participants. For any sites that pool data from multiple sub-awardees, we will weight each sub-awardee proportionate to size, so that each study participant is weighted equally. However, we will also run sensitivity analyses with each sub-awardee given equal weight.

To control for multiple hypothesis testing (the increased chance of falsely identifying an impact as statistically significant when examining effects on many outcomes), we will limit the primary analyses for each site to a small set of key outcomes. In selecting these outcomes, we will rely on the program logic model and data needs table developed for each site. We anticipate that most of these outcomes will be measures of sexual risk behavior and its health consequences (pregnancy, STIs, or birth), though the exact outcomes selected will vary by site. Within this small set of key outcomes, we will also consider applying a formal statistical correction for multiple hypothesis testing.

To support these analyses, the follow-up surveys will include measures of all key outcomes—primarily pregnancy, STIs, and associated sexual risk behaviors. We also included these measures on the baseline survey, so that we can include them as covariates in the regression models used to estimate program impacts.

Exploratory impact analyses of secondary research questions. In addition to our primary impact analysis, we will also define and answer additional 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 provides an estimate of the difference in the program effect across the subgroups. Subgroups of particular interest include gender and baseline sexual experience. To support these analyses, we included these subgroup variables on the baseline survey.
- **Impacts on mediating variables.** In addition to primary analysis of program impacts on outcomes of most central importance, as part of secondary analysis we will also examine program impacts on key mediating variables specified in the program logic model for each site (for example, refusal skills, attitudes, or engagement in after-school or community activities). We will estimate impacts on these outcomes following the same approach described in Equations (1) and (2). These mediating variables will be drawn primarily from the first follow-up survey, which will be conducted 8-12 months after random assignment. We also included selected

mediating variables on the baseline survey, to include as covariates in the regression models.

- ***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 to calculate estimates of the treatment on the treated (TOT). Second, to explore the association between program dosage—the degree of program participation—and impacts, we can conduct propensity score analyses, whereby youth with the highest program attendance are matched to a subset of control group youth with similar demographic and baseline characteristics. To support these analyses, the baseline survey includes a broad range of demographic and other personal characteristics to consider as potential matching variables.

In-depth implementation analysis. The in-depth implementation analysis will feature both quantitative and qualitative analyses. The quantitative analysis will have two main components. First, we will prepare basic descriptive statistics on the characteristics of program participants and program participation rates. This information will be used in generating overall profiles of each site and to describe the target populations. Second, data from the online survey of frontline staff and supervisors will be used to examine the characteristics and attitudes of these staff and to construct quantitative indicators of program fidelity and implementation quality.

For the qualitative analysis, we will systematically code and analyze the data using qualitative data analysis software, such as Atlas.ti or NVivo. First, the lead members of the implementation evaluation team will create a coding scheme consisting of a hierarchy of conceptual categories and classifications linked to the evaluation research questions, dimensions of implementation, and program logic models. A trained team of coders will then use the software to assign codes to specific text in the site visit and interview notes and other documents. Coding the qualitative data in this way will enable the team to access data on a specific topic quickly and to organize information in different ways to make it easier to identify themes and compile the evidence supporting them. As data collection proceeds, the coding scheme will be refined to better align it with themes and topics that emerge from the data and with the research questions (Ritchie and Spencer 2002). To facilitate analyses of patterns and themes across sites, we will also code key site-level characteristics, such as type of program model and characteristics of the youths served.

After all the qualitative data have been coded, we will use the software to retrieve data on the research questions and subtopics to identify themes and triangulate across data sources and individual respondents. Much of the meaning of the data will be discerned through descriptive analyses—qualitative and quantitative—that organizes data thematically. We will create summary statistics that characterize overall experiences in each site, as well as variations across and within sites, and examine themes and topics from multiple perspectives and highlight the similarities and differences among them (Patton 2002). We will also explore relationships across themes (for example, relationships between the types of implementation challenges sites face and their staffing patterns and partnership arrangements).