ATTACHMENT E

ANALYSIS PLAN

ANALYSIS PLAN FOR IMPACT STUDY

The purpose of The Federal Evaluation of Making Proud Choices! (MPC!) is to rigorously assess the impacts of the program in 42 schools. Each participating school will be randomized to one of two conditions – (1) a treatment condition where health educators provided by a local organization deliver MPC!, or (2) a control condition where school health teachers provide their regular health curriculum. Program impacts will be analyzed using survey data collected at baseline and at 9 and 15 months after baseline.

Our analysis plan for the impact study has three main components: (1) an early analysis of baseline data, (2) a primary impact analysis of key behavioral outcome measures, and (3) exploratory analyses of secondary research questions. These are described below.

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. 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. 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 blocking variables used in conducting random assignment or schools (such as the use of a school district as a block), or for any differences between the treatment and control groups in baseline characteristics that arise by chance or from survey nonresponse.

With schools, rather than individual youth, as the unit of assignment, the estimation must account for the correlation of outcomes among youth in the same school, 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, the regression model can be specified as

$$y_{is} = \beta' x_{is} + \lambda T_s + \eta_s + \varepsilon_{is}.$$

In this model, y_{is} is the outcome measure for individual i in cluster s (and similarly for the treatment status indicator T_s, vector of baseline characteristics x_{is} and the error term ε_{is}). Most important, the error term in 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. The estimated impact of the program is λ.

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 to a small set of key outcomes. In selecting these outcomes, we will rely on the program logic model. We anticipate that most of these outcomes will be measures of sexual initiation and sexual risk behavior (such as contraception use). 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 sexual initiation and sexual risk behaviors. We will also include these measures and related measures on the baseline survey, so that we can include them as covariates in the regression models used to estimate program impacts.

Analysis of secondary research questions. In addition to our primary impact analysis, we will also define and answer additional secondary research questions:

- *Subgroup analyses*. To examine whether the programs were more effective for some youth than for others, we may estimate impacts for subgroups of youth by adding a term to the model that interacts the treatment indicator by 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 race/ethnicity, gender, and sexual experience at baseline. To support these analyses, we will include 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 example, knowledge of contraception and attitudes about use of contraception and sexual initiation). We will estimate impacts on these outcomes following the same approach as primary impact analysis. These mediating variables will be drawn primarily from the short-term follow-up survey, which will be conducted 9 months after baseline. We will also include selected mediating variables on the baseline survey, to include as covariates in the regression models.

- *Variation in impacts by quality of program delivery.* Our primary impact analysis will include the full study sample, yielding intent-to-treat (ITT) estimates that do not account for varying quality of programming among youth assigned to the two treatment groups. As exploratory analyses, we will explore the association between program quality—the extent to which the evidence-based program was delivered with fidelity by classroom teachers and outside health educators—and impacts. To accomplish this, we will conduct two separate analyses: (1) Estimate impacts separately by site (e.g. district), and compare the magnitude of the impacts in sites with high levels of fidelity against the impacts in sites with lower levels of fidelity (2) Estimate impacts by comparing individuals who had high attendance and high fidelity of implementation against individuals who did not have high attendance or fidelity of implementation. To support these analyses, we will combine the survey data used to estimate impacts, with the rich implementation data, including attendance, observation, and potentially, interview data, as a means to quantify the quality of delivery as a variable that might explain variation in outcomes/impacts.