FYSB PERSONAL RESPONSIBILITY EDUCATION PROGRAM (PREP): PROMISING YOUTH PROGRAMS (PYP)

INSTRUCTIONS FOR PREIS IMPACT AND PROGRAM IMPLEMENTATION EVALUATION ANALYSIS PLAN

These analysis plan instructions and the accompanying template were designed to help Personal Responsibility Education Innovative Strategies (PREIS) evaluators ultimately produce evaluation reports based on methods likely to receive favorable reviews from systematic reviews and other key stakeholders in the field. Grantees should work with their evaluators to develop a plan that documents the outcomes that will be examined and the approaches to be used to assess program effectiveness. The analysis plan can serve as a tool to think through the necessary decisions for your analysis and ensure that the grantee and evaluator agree with the proposed approach.

An analysis plan, developed a priori, is a way to demonstrate a commitment to objectivity and to a pre-specified, systematic, and scientific approach. The analysis plan is a document that shows funders and potential skeptics the outcomes that were pre-selected (based on a logic model) to show program effectiveness, the analytic approaches to be used to assess program effectiveness, and the justification for those decisions. Pre-specifying an analytic approach can prevent concerns about *data mining*. Data mining refers to selectively choosing results or data, which leads to misleading results. One example is if a researcher runs analyses for dozens of outcomes (which increases the chances of finding a statistically significant effect, by chance) and then only reports on the few outcomes with statistically significant results. Having an analysis plan that clearly describes the analytic approach to be used to estimate program effectiveness provides a road map for undertaking your analysis once data collection is complete and is important should you experience any changes in key project staff.

The analysis plan guidance is organized as follows: Sections 1 and 2 revisit proposed research questions and some aspects of the study design. Section 3 discusses the implementation analysis, and Section 4 discusses how to articulate the benchmark and supplemental analyses used to show program effectiveness. The guidance provides the topics you should cover in your analysis plan and some considerations for your approach or the articulation of your plans. We have created a separate template [ADD LINK] for you to use to draft your plan. As you develop your plan, feel free to reach out to your rigorous evaluation technical assistance (RETA) liaison with any questions or issues you would like to discuss before submitting your plan.

Please email your analysis plan to your Federal Project Officer (FPO) and copy your RETA liaison by [**date**]. For consistency, use this common naming convention when submitting your analysis plan: [**Grantee Name**] Impact Analysis Plan. If you have made substantial changes to your evaluation (for instance, sample size, eligible population, data collection plan) since your last abstract submission that was approved by your FPO, please also submit an updated abstract,

with tracked changes. The latest version of your abstract is available in your grantee SharePoint folder, in the Abstract subfolder. Your FPO and RETA liaison will review the latest version of your abstract, for context, along with your analysis plan, and will provide comments and suggested edits and return to you for revisions. FYSB would like your FPO to be able to approve your analysis plan and updated abstract (if applicable) by [**date**] so that you can begin work on the final report.

1. **Impact study research questions**

   a. **Primary research question(s)**. For the purposes of the analysis plan, primary research questions are those focused on the behavioral outcomes most important to gauging a program's effectiveness in improving adolescent reproductive health. Each primary research question should focus on the effect of the program on a behavioral outcome measure relevant to the HHS TPP Evidence Review at a specific time point. (For more information on the HHS TPP Evidence Review, please see the 2017 webinar materials on SharePoint.) The Evidence Review accepts measures in the following domains: (1) sexual activity, (2) number of sexual partners, (3) contraceptive use, (4) STIs or HIV, and (5) pregnancies. The outcome(s) and the time point(s) should be clearly connected to the program's logic model for the theory of change. For example:

   - "What is the impact of [intervention being tested in evaluation] relative to [counterfactual] on sexual initiation one year after the end of the intervention?"

   - "What is the impact of [intervention being tested in evaluation] relative to [counterfactual] on risky sexual behavior one year after the end of the intervention?"

   Because the likelihood of a false positive—estimating a statistically significant impact when no causal effect exists—increases in some cases with the number of outcomes studied, grantees are encouraged to limit the set of primary research questions to the minimum number that fairly evaluates the program's effectiveness.

   Research questions or outcomes that are not *critical* to evaluating a program's effectiveness in improving adolescent reproductive health but are still important and of interest to the grantee, stakeholders, researchers, and so on may be considered secondary research questions.

   Your primary research questions should have been identified in your original evaluation design or a later resubmission of your evaluation design plan. Note: Any changes to the primary research questions should have been previously discussed and approved by your FPO. Please make sure that the outcomes reflected in your abstract and analysis plan align.

   b. **Secondary research question(s).** Secondary research questions, if distinguished from primary ones before analysis begins, allow researchers to explore a broad range of possible program impacts and mediating factors without increasing the likelihood of false positives among primary research questions. Secondary research questions should include

explorations of other outcomes that might be influenced by the intervention or other justifiable explorations of program effectiveness (for example, whether the program works better for certain populations). For instance:

i. Impacts on outcomes considered pre-cursors or intermediate outcomes to the evaluation's primary behavioral outcomes (for example, self-efficacy for using condoms, knowledge about STIs. These may be precursors to behavioral outcomes like condom use during recent sexual activity, in that youth with higher self-efficacy and more knowledge might be more likely to use condoms in practice).

ii. Impacts on other behavioral outcomes not considered to be the primary, intended outcomes of the program (for example, substance use, impulsive behavior, school attendance)

iii. Impacts for specific subgroups of people (for example, female participants, youth who had not had sex before the enrolling in the study)

iv. Impacts on primary research question outcomes at different time points (such as immediately after the end of the intervention or six months later)

v. Non-experimental exploration of how the program's core components influence adolescents' outcomes [Add LINK to brief once released]

As with the primary research questions, make sure the outcomes in your abstract align with your secondary research questions. Note: All outcomes listed in your abstract should appear under either a primary or secondary research question, or in Section 3 as an implementation research question, within this analysis plan.

2. Impact study design

Briefly describe the study design and the process for creating the intervention (that is, program or treatment) and comparison groups.

a. If a random assignment design: articulate key aspects of the random assignment process

i. Describe the unit of randomization (for example, schools, classrooms, individuals)

ii. Discuss who conducts random assignment, when, and under what circumstances

1. Is randomization conducted by evaluation staff or by program staff?

2. When does random assignment occur with respect to the timing of consent and baseline data collection? For clustered randomized controlled trials (CRCTs), who, if anyone, was told of the outcomes of random assignment before consent and baseline data were collected, and for what purposes?

3. What is the method of random assignment (for example, random number generation in Excel)?

4. Does randomization occur all at once (meaning many units are randomly assigned at a single point in time) or on a rolling basis (that is, small numbers of units are randomly assigned at different points in time)? Describe the details of this process.

(a). Describe any stratification or blocking you use to create separate instances of random assignment in the evaluation (for example, random assignment of students to condition might occur separately across schools—in this situation, schools are strata/blocks). Describe how single units that could not be paired or blocked with others (singletons) are assigned to condition.

(b). If applicable, describe any subsampling that occurred after random assignment, the reason for the subsampling, the criteria used for subsampling, and how the subsampling was operationalized.

(c). Report the intended probability of assignment to the intervention group. If it varies systematically (for example, across blocks/strata), report why and give the range of probabilities used.

(d). Describe any potential opportunities for crossover or contamination during the program. Crossover occurs when people randomly assigned to the intervention or comparison condition are later found to be receiving the services intended to be offered to the other condition. Contamination occurs when people assigned to the comparison condition end up receiving some or all of the conditions intended only as part of the intervention. If a separate, alternative program is part of the comparison condition (as opposed to "business as usual"), describe potential points of contamination for that condition as well. Describe ways in which you monitor, prevent, or minimize crossover and contamination during the evaluation.

b. **If a quasi-experimental design: describe how the research groups were formed**

i. Describe the criteria used to determine whether people (or groups of people) would be assigned to the intervention or the comparison group, and the process used for constructing the intervention and comparison groups. When did this assignment procedure occur, relative to the timing of consent and baseline data collection?

3. **Program implementation analysis**

Describe the implementation research questions and data you will review to understand and document program implementation. Describe any targets you pre-specified and used, if applicable, to assess how well the program was implemented relative to program or developer

standards. At minimum, include measures of the following implementation elements: fidelity, dosage, quality, engagement, and context. Use a table to link the implementation element, research question, measures, and targets. Table 1 presents an example of such a table (sample text included in italics).

Table 1. Planned implementation analysis

Implementation element	Research question	Measure	Target
Fidelity	*Were all intended program components offered and for the expected duration?*	• *Total number of sessions delivered* • *Average session duration, calculated as the average of the recorded session lengths (in minutes)*	• *95 percent of groups to receive all 12 sessions* • *Average session duration will be at least 40 minutes*
Fidelity	*What content did the youth receive?*	• *Total number of topics covered, calculated as the average of the total number of topics checked by each program facilitator in the daily fidelity tracking log or protocol*	• *95 percent of groups to receive 90 percent of the topics*
Fidelity	*Who delivered services to youth?*	• *Number and type of staff delivering services to study participants, such as the number of session facilitators* • *Percentage of staff who receive minimum training, calculated as the number of staff who received at least 20 hours of training divided by the total number of staff who delivered the program*	• *Three full-time health educators will deliver programming* • *All health educators to receive at least 20 hours of training each year*
Fidelity	*What were the unplanned adaptations to key program components?*	• *List of unplanned adaptations, such as a change in setting, sessions added or deleted, and components cut*	• *n/a*
Dosage	*How often did youth participate in the program on average?*	• *Average number (or percentage) of sessions youth attended* • *Percentage of the sample attending the required or recommended proportion of sessions* • *Percentage of the sample that did not attend sessions at all*	• *n/a* • *75 percent of youth to attend 75 percent of the program sessions* • *Less than 5 percent of the sample gets none of the program*
Quality	*What was the quality of staff–participant interactions?*	• *Percentage of observed sessions with high quality interactions, calculated as the percentage of observed interactions that study staff scored as "high quality"*	• *90 percent of observed sessions to be implemented with high quality (rated as a 3.5 out of 4 on the quality scale)*

Implementation element	Research question	Measure	Target
Engagement	*How engaged were youth in the program?*	• *Percentage of observed sessions with moderate participant engagement, calculated as the percentage of sessions in which study staff scored participants' engagement as "moderately engaged" or higher*	• *90 percent of observed sessions to be implemented with moderate to high engagement*
		• *Average engagement rating, calculated as the average of engagement scale scores (ranging from 1–5, for example) across satisfaction surveys*	• *n/a*
Context	*What other pregnancy prevention programming was available to study participants?*	• *Percentage of the sample receiving pregnancy prevention programming from other providers, constructed from immediate post-survey data on experiences outside of the current program*	• *Less than 20 percent of youth to receive formal content outside of the program*
		• *List of pregnancy prevention programming available to study participants outside of the current program, as described on the websites from other agencies in the community*	
Context	*What external events affected implementation?*	• *Percentage and total number of sessions not delivered due to event in the community, if any*	• *n/a*

4. Impact analysis

This section should lay out the specific plans for cleaning your data and handling missing data, constructing your outcomes, defining your analytic sample, assessing baseline equivalence, addressing potential crossover and contamination, and finally, analyzing your data to estimate program impacts and conduct sensitivity analyses.

The analysis plan for evaluating impacts should lay out in advance the outcomes from your research questions and a "benchmark" analysis for the final report. The benchmark analysis is the analytic approach you will use to estimate the findings you will lead with in the summary of findings (that is, the analysis you believe is the most defensible and credible). For instance, you might use a complete case analysis, with no imputed data, as your benchmark approach. You might want to perform additional analyses that alter one or more decisions that informed the benchmark approach to understand how results depend on features chosen for the main analysis. We refer to these subsequent analyses as sensitivity analyses, as they can provide information on the extent to which certain results are sensitive to decision points made for the main analysis. For instance, a sensitivity analysis could be using multiple imputation to impute missing data, relative to the complete case approach used for the benchmark analysis. Sensitivity analyses might be specified in advance or undertaken after uncovering any unforeseen issues such as missing data or problems with covariates or modeling details. (See Selecting Benchmark and Sensitivity Analyses for a discussion of how to approach selecting your benchmark and sensitivity analyses.)

a. **Data cleaning.** Indicate the systems and/or software you use to prepare, clean and store data. Describe the process you will use to identify missing, inconsistent, or inaccurate data, including at what points in the process you review and clean data. Describe in detail how you will then handle missing data, responses that are inconsistent with each other (within a survey or over time), and seemingly inaccurate data, across both baseline and outcome surveys. For example, if a person indicates at baseline that they have ever had sex, but then at a follow-up period indicates they have never had sex, discuss how you will deal with this in your data or analysis. See this resource that discusses how to deal with missing data in RCTs.

b. **Outcome measures.** Describe the specific outcome measures used to answer the primary (and secondary, if applicable) research questions. If the measures will be constructed from different items, include the source item(s) used to create each constructed measure and how you will code it to create the measure.

   i. Complete Table 1, describing all measures you will use to answer the primary research questions assessing the impact of the program. Include the time periods you will use to assess effectiveness.

   ii. Complete Table 2 for all measures you will use to answer secondary research questions. Include the time periods you will use to assess effectiveness. **Finally, please attach the survey instrument(s) as an appendix to the submitted plan.**

Table 2. Behavioral outcomes used for primary research questions

Outcome name	Source item(s)	Constructed measure	Timing of measure
Ever had sexual intercourse	*Have you ever had sexual intercourse*	*Dichotomous variable coded as 1 if answered yes and zero if no and missing otherwise.*	*6 months after program ends*

Table 3. Outcomes used for secondary research questions

Outcome name	Source item(s)	Constructed measure	Timing of measure
Recent sexual intercourse	*Have you had sexual intercourse in the past three months?*	*Dichotomous variable coded as 1 if answered yes and zero if no. Zero if no to "Ever had sexual intercourse" as of the 12-month follow-up. Missing otherwise.*	*12 months after program ends*

c. **Analytic sample(s).** Describe how you will define the analytic sample (for each research question, if applicable). Clearly articulate which data are required for a person to be included in the analytic sample. For example, perhaps the analytic sample for the study will be people with complete baseline and outcome data for all variables of interest at a specific follow-up or across all follow-ups (that is, a complete case sample). Or perhaps, the analytic sample might be people who have complete outcome data but some missing baseline data, which will be imputed (as described above). Note: The HHS TPP Evidence Review will assess attrition separately for each analytic sample, assessing the number of people in the baseline sample for which follow-up was not completed or are missing outcome data relative to the randomly assigned sample. For more information, please review the HHS TPP evidence standards and this research brief on sample attrition.

Note that when creating an analytic sample for a particular time point when there are multiple outcomes to examine (with some item nonresponse across the outcomes), the RETA team recommends identifying a single, common analytic sample that does not have missing data across the outcomes of interest. Using a single, common analytic sample will produce an easy-to follow and understandable presentation of the analyses across multiple outcome measures. If, however, there is substantial item nonresponse across two or more outcomes, then we recommend considering each outcome as requiring its own, unique analytic sample (you will need to demonstrate baseline equivalence separately for each analytic sample).

d. **Assessment of baseline equivalence.** What measures will you use to examine the equivalence of the groups at baseline? What methods will you use to test the significance of the difference between the groups? At a minimum, include the demographic and

behavioral measures of interest to the HHS TPP Evidence Review (age/grade, gender, race/ethnicity), as well as baseline measures of each outcome. How will the benchmark approach to impact analyses adjust for any significant differences in baseline measures between groups? For more information, please review the HHS TPP evidence standards and [this research brief on baseline equivalence and matching.](#)

e. **Condition crossover and contamination.** Describe how you will quantify and report the amount of crossover and levels of contamination that occur over the length of the program.

f. **Analytic approach for primary research questions.** Describe how you will conduct the analysis to answer the primary research questions, under an intent-to-treat framework. That is, describe your plans to include all the study participants who were assigned (randomly, if study is an RCT) to the study groups (treatment and comparison) in the impact analysis and examine them in the groups to which you had originally assigned them. In addition, describe your plans to include all participants who provide outcome data (that is, participate in the follow-up data collection) in the impact analysis, even if they do not complete services.

i. **Model specification.** Provide the type of model you will use to estimate program impacts for each primary and secondary research question (e.g., linear regression, logistic regression or MANOVA[1]). For RCTs, we recommend linear probability models because they are easy to interpret. See [this brief for more detail on the rationale](#).

   1. What statistical software package, including the version, will you use?

   2. Define the criteria you will use to assess the statistical significance of study findings (for purpose of the HHS TPP Evidence Review, findings are considered statistically significant based on $p < .05$, two-tailed test).

   3. How will the model adjust for clustering (if applicable)? See [this resource for some frequently asked questions about clustering in RCTs](#).

ii. **Covariates.** List all potential covariates (control variables in the regression) you will include in the analysis in Table 3 and justify your reason for their inclusion. We assume most grantees' benchmark analysis model will include covariates for the outcome measures at baseline and demographic characteristics (including age, gender, race/ethnicity, as applicable), which might enhance the precision of impact estimates. If these are not included in the benchmark model, we recommend at minimum including any baseline measures of the outcome for which you find significant differences between condition groups during the assessment of baseline equivalence. (Note: This is necessary

---

[1] If you choose to use a MANOVA, we recommend you also estimate and report impacts using a linear or logistic regression to facilitate interpretation. MANOVAs are useful for examining the effect of a variable on multiple dependent variables, essentially building in a multiple comparisons correction. The MANOVA results, however, do not provide an easily interpretable metric for stating the effect of the intervention on an individual outcome.

to achieve the highest rating by the Evidence Review if the study design is a RCT. If the study design is quasi-experimental, then the primary analyses (primary research questions) must include the baseline measure of the corresponding outcome to achieve a rating of moderate by the Evidence Review.)

1. If you have not yet determined covariates, describe a plan for determining what covariates you will include. Aside from the baseline version of the outcome of interest, will any covariates differ across the models used to answer the primary research questions? When appropriate, describe how you will incorporate blocking or stratification variables as covariates.

Table 4. Covariates included in impact analyses

Covariate	Description of the covariate
Age	*Age (in years) as of the baseline data collection*

i. **Adjustments for multiple comparisons (if applicable).** Describe the approach you will use to adjust for the multiple hypothesis tests if the study will address more than one primary research question. It is good practice to minimize the occurrence of false positives by adjusting statistical significance tests for the number of comparisons associated with primary research questions. These adjustments should appropriately raise the threshold of statistical significance of impact estimates for each outcome of interest as the number of outcomes increases. If you will conduct multiple hypothesis tests yet will not be using multiple comparison adjustments, justify why you will not be performing such adjustments.

ii. **Sensitivity analyses.** Describe any analyses you will conduct to test the robustness of the results or the appropriateness of the analytic model for the observed data. Include analyses that make variations of potentially important research decisions, such as the procedures used to prepare and handle missing and inconsistent data, the choice of baseline covariates to adjust for stratification or blocking, and so on.

g. **Analytic approach for secondary research questions.** Describe the analytic approach you will use to address all secondary research questions to the extent that it differs from above (for instance, you will not conduct multiple comparison adjustments for secondary research questions). Please cover items in Sections 3.f.i.–3.f.iv.

5. Additional planned analyses

Identify any additional research questions that you plan to address using data from this evaluation, if not mentioned previously. In addition, this section can include alternate specifications used to test impacts of the intervention across time points, such as growth-curve analyses.