The information in this document was produced under contract number (HHSP233201300416G) for the Office of Adolescent Health (OAH), Department of Health and Human Services (HHS).

MATHEMATICA
Policy Research

Final Evaluation Reporting
OAH Tier 1 C/D, Tier 2

Webinar Presentation to Grantees

October 28, 2014

M. C. Bradley

Overview of Presentation

- **Final evaluation report requirement**
- **Discuss changes since June draft**
- **Familiarize audience with template**
- **Questions and Answers about template**

Final Evaluation Report Requirement

- **Separate from grant wrap-up reporting requirement**
- **Presentation of study impacts, with supporting documentation**
  - **20 page limit for body of text**
  - **Supplemental appendices, as needed**
- **Will be publicly available and submitted to the HHS evidence review**
- **Not the only venue for disseminating evaluation findings**

Perspective

- **Write for naïve audience, not liaison or project officer.**
- **Orient reader to YOUR study and how it fits in the larger effort.**
- **Tell a story that is easy to follow.**
  - **What were you looking at?**
  - **What programming was intended? Provided?**
  - **How did you study the impacts?**
  - **What did you find?**
- **Focus on benchmark analysis**

Benchmark versus Sensitivity Analyses

- **A benchmark analysis is the key or central analysis in a report.**
  - **For the OAH report, should be the analysis that is of interest to the HHS Evidence Review**
- **Sensitivity analyses are supplemental analyses that change one aspect of the benchmark analysis.**
  - **Examples include:**
    - Intent-to-Treat analysis for an RCT with high attrition that presented a benchmark analysis that matched intervention and comparison participants
    - An analysis that imputes outcome or covariate data.
    - An analysis that uses an alternate matching strategy.

Changes from Draft (June 2014) to Final

- **Added table shells and instructions.**
- **Added multiple comparison adjustment guidance.**
- **Added reminders that analyses with imputed data should not be used for benchmark analyses for RCTs with high attrition or QEDs.**
- **Expanded appendices to include additional, optional table shells for supplemental detail and sensitivity analyses.**
- **Changed location of the intent-to-treat (or ITT) analyses for RCTs with high attrition.**
  - **The matching analysis should appear in the main body. The ITT analysis would be presented in the appendix as a sensitivity analysis.**

Example

I. Introduction (approximately 2 pages)

A. Introduction and Study Overview

Purpose	Orient the reader to the study.
Instructions and Reminders	In this section, explain (1) the need for teen pregnancy prevention for the particular population (defined by locality, age, gender, race/ethnicity, etc.) studied; (2) the program selected and how the evaluation fit within the OAH/FYSB grant program—in particular, the funding tier (1 or 2) used for this project; (3) that this report describes the implementation and impact of the TPP-funded program; and (4) previous research describing the effects of the program, including, if applicable, how prior findings were assessed by the HHS evidence review. The reader should understand why the program was targeted to certain youth and the motivation for selecting the chosen program.
Potential Sources	2010 Funding Opportunity Announcement for Tier 1 and Tier 2/PREIS Grant proposal
Non-Text Elements	None.

B. Primary Research Question(s)

Purpose	Primary research questions articulate the main confirmatory hypotheses about behavioral outcomes that are tested in a TPP impact evaluation.
Instructions and Reminders	This section should present the primary research questions. Reminder: The primary research question(s) should focus on the impact of the program on at least one behavioral outcome measure that is relevant to the HHS Pregnancy Prevention Evidence Review. All primary research questions should also focus on impacts at a specific time point. The outcome(s) and time point(s) should be clearly connected to the theory of change for the program.
Potential Sources	Impact Analysis Plan: Section 1 Evaluation Abstract
Non-Text Elements	None.

Example

F. Baseline Equivalence

Purpose	Provide information on how baseline equivalence was assessed for the analytic sample(s), and present the results of the analysis assessing the equivalence of the analytic samples used to answer primary and secondary research questions.
Instructions and Reminders	Briefly describe the analytic methods used to assess the equivalence of the analytic sample(s). Reminder: the analytic method used to show baseline equivalence should account for the study design (for example, clustering, stratification). Equation(s) for estimating equivalence of analytic sample(s) can be included in the appendix if necessary. Present an equivalence table for each analytic sample being used to answer the primary and secondary research questions. For example, if the primary research questions focus on a 12-month follow-up assessment, and the secondary research questions focus on a 24-month follow-up assessment, provide tables for: (1) the sample responding to the 12-month follow-up with nonmissing data on recent sexual activity (primary analytic sample), and (2) the sample responding to the 24-month follow-up (secondary analytic sample). The baseline equivalence tables must include demographic characteristics (age or grade, gender, and race/ethnicity), as well as prior measures (or highly correlated measures) of the outcomes. For each group, the table should document (1) sample sizes for each characteristic reported, (2) mean and standard deviation for continuous variables, and (3) the proportion, as a decimal for categorical variables.
Potential Sources	Impact Analysis Plan: Section 4b Frequently Asked Questions Document for Impact Analysis Plan Biannual reporting tables
Non-Text Elements	The following page includes a (landscape) table shell for Table III.3 to be used to demonstrate baseline equivalence. Baseline equivalence tables are required.

Example: Table Instructions

Non-Text Elements	The following page includes a (landscape) table shell for Table III.3 to be used to demonstrate baseline equivalence. Baseline equivalence tables are required. **Instructions for Completing Table III.3** • The purpose of this table is to demonstrate equivalence between groups on key baseline characteristics. • Copy and paste this table so there is one table for each analytic sample in the report. An analytic sample is described as the sample on which effects are estimated. For example, to report on two analytic samples (6-month follow-up and 12-month follow-up), provide two baseline equivalence tables. • Replace the "[Survey Name]" text with the time point of the survey. For example, "Table III.3. Summary Statistics of Key Baseline Measures for Youth Completing the 6-month Follow-Up Survey." • Replace the "Behavioral measure X" text with the name of the behavioral measure(s) for which baseline equivalence is assessed. • Please add rows for additional measures, as needed. If the sample members are young and did not complete the baseline measure of the behavioral outcome, please report equivalence on the variables collected at baseline that might be correlated with outcomes (if available). • In columns 2 and 3 ("Intervention" and "Comparison"), enter the mean and standard deviation (or percentage) for the baseline measure noted. • In columns 2 and 3, if the characteristic is a continuous variable, enter the mean value and the standard deviation. If the characteristic is binary (or dichotomous), enter the percentage as a decimal (that is, 0.50 instead of 50% if 50% of the sample was female).

Example: Table Shell

Table III.3. Summary statistics of key baseline measures for youth completing [Survey Name]

	Intervention	Comparison	Intervention versus comparison	
Baseline measure	Mean or % (standard deviation)	Mean or % (standard deviation)	Mean difference	*p*-value of difference
Age or grade level				
Gender (female)				
Race/ethnicity				
White				
Black				
Hispanic				
Asian				
Behavioral measure 1				
Behavioral measure 2				
Sample size				

Appendices

- **Appendices should be incorporated, as needed, to tell the story.**
- **Seven possible appendices were identified, and templates generated for five of the seven.**
  - **Logic Model**
  - **Data Collection Timing (Appendix A)**
  - **Implementation Data Collection (Appendix B)**
  - **Sample Flow (Appendix C – REQUIRED)**
  - **Implementation Data Methods (Appendix D)**
  - **Sensitivity Analyses (Appendix E – including ITT estimates for RCT with high attrition)**
  - **Model specifications for baseline equivalence and/or program impacts**
  - **Methods to clean and prepare data (including descriptions of how missing data and inconsistent data were handled)**
  - **Detailed descriptions of methods used to analyze implementation data**

Appendix C: Study Sample

Table C.1a. Cluster and youth sample sizes by intervention status – cluster designs

	Time period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Number of Clusters							
1. At beginning of study		*1 (=1a + 1b)*	*1a*	*1b*			
2. Contributed at least one youth at baseline	*Baseline*	*2 (=2a + 2b)*	*2a*	*2b*	*=2/1*	*=2a/1a*	*=2b/1b*
3. Contributed at least one youth at follow-up	*Immediately post-programming*	*3 (=3a + 3b)*	*3a*	*3b*	*=3/1*	*=3a/1a*	*=3b/1b*
4. Contributed at least one youth at follow-up	*6-months post-programming*	*4 (=4a + 4b)*	*4a*	*4b*	*=4/1*	*=4a/1a*	*=4b/1b*
5. Contributed at least one youth at follow-up	*12-months post-programming*	*5 (=5a + 5b)*	*5a*	*5b*	*=5/1*	*=5a/1a*	*=5b/1b*
Number of Youth							
6. In non-attriting clusters/sites at time of assignment		*6 (=6a + 6b)*	*6a*	*6b*			
7. Who consented		*7 (=7a + 7b)*	*7a*	*7b*	*=7/6*	*=7a/6a*	*=7b/6b*
8. Contributed a baseline survey		*8 (=8a + 8b)*	*8a*	*8b*	*=8/6*	*=8a/6a*	*=8b/6b*
9. Contributed a follow-up survey	*Immediately post-programming*	*9 (=9a + 9b)*	*9a*	*9b*	*=9/6*	*=9a/6a*	*=9b/6b*
10, Contributed a follow-up survey	*6-months post-programming*	*10 (=10a + 10b)*	*10a*	*10b*	*=10/6*	*=10a/6a*	*=10b/6b*
11. Contributed a follow-up survey	*12-months post-programming*	*11 (=11a + 11b)*	*11a*	*11b*	*=11/6*	*=11a/6a*	*=11b/6b*

Sample Appendix C

Table C.1a. Cluster and youth sample sizes by intervention status – cluster designs

	Time period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Number of Clusters							
1. At beginning of study		*40*	*20*	*20*			
2. Contributed at least one youth at baseline	*Baseline*	*40*	*20*	*20*	*100*	*100*	*100*
3. Contributed at least one youth at follow-up	*Immediately post-programming*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
4. Contributed at least one youth at follow-up	*6-months post-programming*	*39*	*19a*	*20*	*97.5*	*95*	*100*
5. Contributed at least one youth at follow-up	*12-months post-programming*	*35*	*19*	*16b*	*87.5*	*95*	*80*
Number of Youth							
6. In non-attriting clusters/sites at time of assignment		*4,257*	*2,218*	*2,039*			
7. Who consented		*3,907*	*2,018*	*1,889*	*91.7*	*90.9*	*93.6*
8. Contributed a baseline survey		*3,179*	*2,000*	*1,779*	*74.6*	*90.2*	*87.2*
9. Contributed a follow-up survey	*Immediately post-programming*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
10, Contributed a follow-up survey	*6-months post-programming*	*3,790*	*1,990*	*1,800*	*89.0*	*89.7*	*88.3*
11. Contributed a follow-up survey	*12-months post-programming*	*3,500*	*1,800*	*1,700*	*82.2*	*81.2*	*83.4*

The full example Table C.1a and related consort diagram will be posted on Sharepoint with this presentation.

Expectations

- **Each grantee will submit a draft evaluation report, based on the provided template, for review by Mathematica.**
  - **Due date determined by federal project officer and in NCE letter**
- **Reach out to liaison and project officer if realize need to do something differently than as documented in analysis plan or NCE letter**
  - **Will require written documentation of changes along with final report**
- **Mathematica will review and provide feedback on evaluation report**
  - **Anticipate 3 – 6 months for review process**
  - **At least 2 reviewers with one not previously involved (cold-read)**
  - **Comments may address any aspect of the evaluation report**
  - **Similar to review of impact evaluation analysis plans**

Next Steps

- **Liaison has been confirming final report dates and content with OAH grantees this month.**
- **Encourage you to begin writing sections before data collection is complete (e.g. Introduction, Program and Comparison Programming, some sections of Study Design).**
- **Encourage you to clean data, build analytic models, and run models with data in hand, particularly if small window of time between when your data collection ends and your report is due.**
- **Keep project officer and liaison informed about progress, questions, and potential changes to impact or implementation analysis plans.**

FAQ

- **Can we adjust page lengths if the overall is still 20 pages?**
  - Yes, the section page lengths are suggestions although the 20 pages is definite.
- **Does the title need to be "Impact Evaluation of [Grantee Program Name] in [Location]"?**
  - No, you may use a title that resonates for you.
- **Do we need to do all seven appendices?**
  - No, not all grantees will need all appendices. As indicated on Slide 10, only Appendix "C" is required.
- **Do we need to demonstrate baseline equivalence even if there is low attrition at the unit of assignment?**
  - Yes, although it is not required to meet HHS evidence review standards, reporting on equivalence will help the reader understand the sample as chance differences can occur. Also, if there are statistically significant differences, the analysis include controls for those variables.

Questions?

For More Information

- Amy Farb
  Amy.Farb@hhs.gov
- Jean Knab
  JKnab@mathematica-mpr.com
- Russ Cole
  RCole@mathematica-mpr.com
- Cay Bradley
  CBradley@mathematica-mpr.com