

Design and Analysis Plan
Quantitative Synthesis of Federally-Funded Teen Pregnancy Prevention Programs
HHS Contract #HHSP233201500069I
5/2/2016

Overview

The goal of the meta-analysis is to assess the effects of teenage pregnancy prevention (TPP) programs on sexual behavior outcomes, with specific emphasis on examining whether program effects vary according to various aspects of program design, program implementation, participant characteristics, and study methods. The meta-analysis will also examine whether program characteristics affect participant retention in TPP programs. A random-effects meta-regression modeling framework will be used to address each of the project's research questions, where effect sizes indexing program effects will be regressed on a range of effect size moderators (i.e., study characteristics) related to program design, program implementation, participant characteristics, and study methods. In the event that multiple, statistically dependent effect sizes are available from each study, the meta-regression models will be fitted using robust variance estimates. Finally, individual participant data will be used (again, under a random-effects meta-regression modeling framework) to provide more refined examination of how participant characteristics are associated with program effects. Below, we provide more detailed description of the proposed analytic strategies.

Statistical Procedures

Effect size metric

We will compute effect sizes for all reported outcome measures in three domains: sexual activity (e.g., abstinence, unprotected sex, number of sexual partners, number of sexual experiences), sexually transmitted infections (e.g., any STI, number of STIs), and pregnancy/births (e.g., any pregnancy, number of pregnancies, any births). All analyses will be conducted separately for these three outcome domains.

Standardized mean difference effect sizes will be used as the effect size metric (Lipsey & Wilson, 2001). All effect sizes will be coded so that positive effect sizes represent better outcomes (e.g., more abstinence, less unprotected sex) for the group receiving the intervention. Standardized mean difference effect sizes (d) are calculated as:

$$d = \frac{\bar{x}_{TX} - \bar{x}_{CT}}{s_{pooled}}$$

where the numerator is the difference in group means for the intervention group (\bar{x}_{TX}) and comparison group (\bar{x}_{CT}), and the denominator s_{pooled} is the pooled standard deviation for the intervention and comparison groups. All standardized mean difference effect sizes will be adjusted with the small-sample correction factor to provide unbiased estimates of effect size (g) (Hedges, 1981). The small-sample corrected effect size g and its standard error are calculated as:

$$g = \left[1 - \left(\frac{3}{4N-9} \right) \right] * d$$

$$SE_g = \sqrt{\frac{n_{TX} + n_{CT}}{n_{TX} * n_{CT}} + \frac{g^2}{2(n_{TX} + n_{CT})}}$$

where N is the total sample size for the intervention and comparison groups, d is the original standardized mean difference effect size, n_{TX} is the sample size for the intervention group, and n_{CT} is the sample size for the comparison group.

For binary outcomes, the Cox transformation outlined by Sánchez-Meca and colleagues (2003) will be used to convert log odds ratio effect sizes into standardized mean difference effect sizes:

$$g_{Cox} = \frac{\ln\left(\frac{A * D}{B * C}\right)}{1.65}$$

$$SE_{g_{Cox}} = 0.367 \left[\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D} \right]$$

where A is the count of “successes” in the intervention group (e.g., number of participants who were abstinent), B is the count of “failures” in the intervention group (e.g., number of participants who were not abstinent), C is the count of “successes” in the comparison group, and D is the count of “failures” in the comparison group. Because the Cox transformation assumes

the binary outcome measures are from an underlying continuous distribution, all analytic models will include (i.e., adjust for) a dummy variable indicating whether the effect size was calculated via the Cox transformation for dichotomous outcomes. If sensitivity analysis indicates this variable has a substantial impact on model results, alternative strategies will be used to explore whether separate analyses should be conducted on the subset of binary outcomes using log odds ratio effect sizes.

In the event that most calculated effect sizes are for binary outcomes, and only a small subset of standardized mean difference effect sizes emerge for continuous outcomes, we will instead use the Cox transformation to convert all standardized mean difference effect sizes into a log odds ratio effect size:

$$LOR_{Cox} = g * 1.65$$

Again, all analytic models will include (i.e., adjust for) a dummy variable indicating whether the effect size was calculated via the Cox transformation for continuous outcomes. Again, if sensitivity analysis indicates this variable has a substantial impact on model results, we will explore alternative strategies that separate all analyses for the continuous and binary outcomes.

Effect size moderators

As detailed in the coding manual for the meta-analysis, a wide range of study characteristics will be coded from the original study reports for descriptive purposes and for use as covariate controls in the final meta-regression models. However, the key effect size moderators of interest will be those related to the program design, program implementation, participant characteristics, and study methods outlined in the project research questions. Each of the research questions for the project is associated with a block of meta-regression models, which will be described in greater detail below.

The first research question focuses on whether program design affects the impact of TPP programs. We will examine six key moderators related to program design: program focus, program components, program teaching/communication strategies, frequency of program contact, duration of the program, and group composition.

- *Program focus* will be measured with a series of binary dummy variables indicating whether the primary focus of the program was: abstinence, sexual health, youth development, HIV/AIDS prevention, or reproductive health services.
- *Program components* will be measured with a series of binary dummy variables indicating whether the program included any of the following components: condom demonstration, service learning, role-plays, games, reflective exercises, mentoring/tutoring, individualized counseling, direct provision of reproductive health or other health services, parent activities, and community outreach.
- *Program strategies* will be measured with a series of binary dummy variables indicating the standard format of the program delivery: individual, small group (<10) with provider, large group or whole classrooms with provider, online, or other.
- *Frequency of contact* will be measured with a series of binary dummy variables indicating whether the program is delivered daily, 3-4 times per week, 1-2 times per week, less than weekly, or one day only.
- *Contact hours* will be measured with a continuous variable indicating the intended length/intensity of the intervention as measured in total hours of contact time.
- *Program duration* will be measured with a continuous variable indicating the number of weeks from first to last contact with program participants.
- *Group composition* will be measured with a binary dummy variable indicating whether the program was delivered to same-sex groups versus mixed- sex groups.

The second research question focuses on whether program implementation affects the impact of TPP programs. We will examine two key moderators related to program implementation: program setting and program delivery personnel. If sufficient data can be collected on the level of preparation or training required for program staff, this research question will be revised to include an indicator of personnel preparation/training as a third moderator.

- *Program setting* will be measured with a series of binary dummy variables indicating whether the intervention was typically delivered in: classrooms, health clinics, community, or other.
- *Program delivery personnel* will be measured with a series of binary dummy variables indicating whether the intervention was typically delivered by: medical professionals,

health educators, classroom teachers, peer educators, other, or mixed (no predominant provider type).

The third research question focuses on whether participant characteristics affect the impact of TPP programs. We will examine four key moderators related to participants: sex, race/ethnicity, age, and prior sexual experience.

- *Participant sex* will be measured with a continuous variable indicating the proportion of males present in the intervention group.
- *Participant race/ethnicity* will be measured with several continuous variables indicating the proportion of White, Black, and Hispanic participants present in the intervention group.
- *Participant age* will be measured with a continuous variable indicating the average age of participants in the intervention group.
- *Participant prior sexual experience* will be measured with a continuous variable indicating the proportion of participants in the intervention group who reported ever having had sex at baseline.

The fourth research question focuses on whether study design, methods, and procedures affect the impact of TPP programs. We will examine six key moderators related to study methods: study design, overall attrition, differential attrition, risk of bias due to sequence generation, intent-to-treat analysis, and handling of missing data.

- *Study design* will be measured with a series of binary dummy variables indicating the study design: individual level random assignment, cluster level random assignment, quasi-random assignment, non-random assignment with matching, or non-random assignment with statistical controls. These categories may ultimately be collapsed into randomized and quasi-experimental design categories.
- *Overall attrition* will be measured with a continuous variable indicating the overall attrition rate in the sample as a whole.
- *Differential attrition* will be measured with a continuous variable indicating the differential attrition rate between the intervention and comparison conditions.

- *Sequence generation risk of bias* will be measured with a nominal three-category measure of low risk of bias, high risk of bias, or unclear risk of bias, using the Cochrane collaboration's item on risk of selection bias due to inadequate generation of a randomized sequence.
- *Intent-to-treat analysis* will be measured with a binary dummy variable indicating whether the authors explicitly reported using an intent-to-treat analysis in their final outcome analyses.
- *Missing data handling* will be measured with a series of binary dummy variables indicating the methods that the authors used to handle missing data: listwise deletion, pairwise deletion, mean/mode imputation, single regression imputation, dummy variable imputation, multiple imputation, FIML, other method, not applicable, or cannot tell.

Analytic Strategies

All analyses will be conducted in a meta-regression framework with robust variance estimates (RVE), which can be used to synthesize statistically dependent effect sizes. Because we anticipate each study sample to provide multiple (dependent) effect size estimates (e.g., for different operationalizations of the same outcome, for multiple follow-ups), the RVE meta-regression model can be used to synthesize all available effect sizes without loss of information.

The RVE meta-regression is similar in form to traditional meta-regression, namely:

$$y_{ij} = \beta_0 + \beta_1 x_{1ji} + \dots + \beta_p x_{pji} + u_j + e_{ij}$$

where for $i = 1 \dots k_j, j = 1 \dots m$, y_{ij} is the i th effect size in the j th study; β_0 is the average population effect; u_j is the study level random effect such that $\text{Var}(u_j) = \tau^2$ is the between-study variance component; and e_{ij} is the residual for the i th effect size in the j th study. In an unconditional meta-regression model, the intercept estimates the traditional average effect size. The above regression equation includes main effects of study-level covariates but may also include multiplicative terms (interaction effects). For instance, we can use multiplicative interaction terms to examine whether different program types have larger or smaller effects on different types of outcomes within the same outcome domain (e.g., do abstinence-only programs have larger effects on abstinence relative to unprotected sex).

We will use a series of variables (described above) to measure the key intervention, participant, and methodological moderators of interest. In addition to these key moderators, the meta-regression models may also include additional covariate controls to address any potential confounding effects. This might include, for instance, information about the strength of the counterfactual condition (e.g., the proportion of controls receiving information within the past year on relationships and dating, reproduction, abstinence, how to say no to sex, STDs, and birth control methods). Because the study samples will span the period of adolescence, the effect of age will be examined closely. Specifically, exploratory analyses will examine whether the effects of the other participant characteristics are moderated by age, and whether there is enough heterogeneity in effects to justify estimating separate models for different age categories.

Differences associated with prior sexual experience will also receive attention to determine if separate models are needed for pregnancy prevention interventions that target youth with little or no sexual history and those that target youth with more sexual experience at baseline. For studies with multiple follow-up measurements, we will code the time from end of treatment as a predictor and use techniques that account for the statistical dependencies for effect sizes from the same sample. This allows for formal testing of the extent to which effects were sustained.

Before conducting the final analysis, we will conduct an outlier analysis to identify any effect size and/or sample size outliers with the potential to distort the analysis; these will be Winsorized to the corresponding lower/upper fence values of their respective distributions and a sensitivity analysis will be conducted to ensure that such decisions do not have an inordinate effect on the findings. These outlier adjustments ensure that such studies will not exercise a highly disproportionate provided to evaluators, we anticipate that a small number of studies may have missing data on method, participant, or treatment variables used in the final analyses. In such cases we will contact the authors to see if they can provide the missing data; any remaining missing values will be imputed using an expectation-maximization (EM) algorithm.

Consistent with standard meta-analysis approaches, we will give more weight to studies whose effect size estimates have greater precision, where precision is primarily a function of study sample size. In the proposed RVE meta-regression approach, the weights include a within-study

as well as a between-study component to the variance. The within-study component is the average variance across effect sizes within the study, and the between-study component is calculated using a method of moments estimator (Hedges, Tipton, & Johnson, 2010).

Individual Participant Data (IPD) Meta-analysis

If raw data on participant-level outcomes can be obtained from some or all evaluators, we will supplement our analyses of the aggregate data (AD) meta-analysis with individual participant data (IPD) meta-analysis approaches. Specifically, we will use a combination of IPD (as available) with AD meta-analysis to examine variability in effects across participant characteristics (i.e., sex, race/ethnicity, age, prior sexual history). Whereas the standard aggregate-data meta-analysis approach is useful for identifying whether study-level moderators (e.g., average age of sample) are associated with larger or smaller program effects, IPD meta-analysis can be useful for examining whether participant-level covariates (e.g., participant age) are associated with program effects. IPD meta-analysis can thus provide more detailed information about program effects for clinically relevant subgroups. If available, we will also examine program effects on other non-behavior outcome measures (e.g., attitudes, intentions).

Because we anticipate that some evaluators will not provide IPD for the project, we anticipate having a mixture of IPD and AD available for synthesis. We will therefore use the one-step approach outlined by Riley et al. (2008) to synthesize findings with a combination of IPD and AD. This method uses a multilevel model to examine the effects of individual participant covariates on program effects, but includes a dummy variable to distinguish between IPD trials and AD trials. In this multi-level model, only the IPD trials contribute information to the analysis examining the effect of participant level moderators, but both the IPD and AD trials contribute information to the overall average program effect as well as the between-study variance component.