Teen Health Study: Analysis Plan

U.S. Department of Health and Human Services
Office of Adolescent Health
Practice Self-Regulation

January 2019

The Policy & Research Group
8434 Oak Street
New Orleans, LA 70118
www.policyandresearch.com
504.865.1545

INTRODUCTION: STUDY CHANGES TIMELINE

OVERVIEW

On April 18th, 2019, The Policy & Research Group reviewed the registry page for the *Evaluation of Practice Self-Regulation* on clinicaltrials.gov and found several inaccuracies in language and content. Updates were made to the registry to reflect the study design documented in our Evaluation Abstract (see Appendix B),[1] submitted to the Office of Adolescent Health on September 17, 2017 and the Impact Analysis Plan (included below), submitted to OAH on January 31, 2019.

In an effort to be transparent about changes that have been made to the registry since its inception, we outline below substantive changes that have been made to the design of the study over the course of the implementation period. Changes are organized chronologically. Detail is provided on what the original content of the Clinical Trials registration included, what the revised content now indicates, and (when applicable) provides a rationale for the change.

MARCH 2016

- The original eligibility requirement of having three or more Adverse Childhood Experiences was removed after the pilot was completed and before the full study began. When reviewing eligibility screening information for pilot enrollees into the study, PRG found that a number of participants were not meeting the ACE criterion. The program developer and clinicians involved in the study suggested that youth might be underreporting ACEs prior to study enrollment because they had not yet developed a relationship with the clinician providing them care at that time. The ACE criterion was ultimately removed after the program developer and trained clinicians working on the study indicated the intervention would be appropriate for youth even if they did not have a score of three or more.

JULY 2017

- Office of Adolescent Health (OAH) Teen Pregnancy Prevention 9TPP) Tier 2B grantees receive notice that the funding period for the grant has been shortened by two years, shifting the end date of the grant from June 30, 2020 to June 30, 2018

APRIL 2018

- Federal judge rules in favor of the OAH TPP Tier 2B grantees and funding is reinstated. Funding period will now end on June 30, 2020.

NOVEMBER 2018

- Anticipated enrollment was modified from 600 participants to 400 participants to reflect current enrollment trends and expected study sample size

[1] Note that several details related to study implementation described in the September 2017 Evaluation Abstract have changed over the course of the study, but for the purposes of transparency and documentation we have included the original abstract in Appendix B.

1) Research Questions that Address Program Effectiveness on Behavioral Outcomes

a. Primary research questions

1. What is the impact of the offer to participate in *Practice Self-Regulation* (treatment) relative to the offer to participate in typical therapy or counseling (control) on participants' reported number of sexual partners nine months after receiving the treatment?

2. What is the impact of the offer to participate in *Practice Self-Regulation* (treatment) relative to the offer to participate in typical therapy or counseling (control) on participants' reported number of times having sex without a condom nine months after receiving the treatment?

b. Exploratory research questions

Exploratory research questions will investigate mediating factors, subgroup effects, outcomes immediately post therapy, and other behavioral outcomes.

2) Description of the Intervention and Counterfactual Condition

The Teen Health Study (THS) is a randomized controlled trial (RCT) in which eligible, consenting participants are randomly assigned to a treatment or control intervention at partnering outpatient mental health agencies, clinics, or private practitioners' offices located in five states – California, Louisiana, Maine, Michigan, and New Mexico.

The intervention is offered to youth, aged 14 to 19, who are receiving outpatient counseling services from a master's-level trained therapist at one of the study's implementation sites, have been deemed appropriate for the study by clinic staff with regards to physical and mental health capacity, have not previously used the *Trauma Outcome Process* (TOP) workbook in therapy, do not live with any youth who are enrolled in the Teen Health Study, have not participated in other Office of Adolescent Health-funded Teen Pregnancy Prevention (OAH TPP) programs, consent to participate in the study, and are willing to complete post-program and nine-month follow-up questionnaires.

The treatment condition, *Practice Self-Regulation* (PS-R), is a manualized therapy intervention that is trauma-focused and conducted one-on-one with youth who are receiving individual outpatient counseling services. It aims to increase knowledge of sexual health and the impact of trauma on decision-making, readiness to change risky sexual behavior, ability to manage impulsive behavior, confidence in one's ability to negotiate safe sex, and intentions to practice safe sex. It also enables participants to explore the effect of previous trauma on their behavior more generally. The primary behavioral goals of the intervention are to reduce unprotected sexual activity and sexual activity with multiple partners.

The control (counterfactual) condition is business as usual; that is, the typical therapy or counseling that participants would receive from their therapists. Therapy models typically used by therapists vary, but other trauma-informed interventions or therapies are used with control participants (e.g., Trauma-Focused Cognitive Behavioral Therapy). Therapists have been instructed that they can answer specific questions related to sexual health in therapy sessions, however, they are to provide information only and referrals for additional services when necessary, as they would under typical circumstances. Partner agency representatives and participating private practitioners have confirmed that no other curriculum-based teen pregnancy prevention programs or similar therapy-based sexual education programs will occur during the study period.
.

a. **Intervention condition:** The intervention condition, PS-R, is a trauma-informed therapy model that is designed to promote healthy adolescent sexual health. It was developed by Joann Schladale of Resources for Resolving Violence, Inc. and is predicated on affect regulation. Affect regulation is a persons' ability to manage emotions without causing harm to self or others.[2] Affect regulation has five core components: thoughts (cognition), feelings (affect), physiological reactions, behavior, and outcomes. The first three are internal processes people learn to mindfully observe as they experience external stimuli (i.e. arousal, but not necessarily sexual). These three internal elements influence decision making that in turn influences behavior and outcomes. PS-R was developed specifically for adolescents, and aims to help young people who have been affected by trauma to recognize and restructure their thoughts and regulate emotion and physiological reactions, with the ultimate goal of helping youth make optimal sexual health decisions.[3]

i. **Intended program components**: PS-R is an empirically-based, trauma-informed therapeutic intervention designed for youth impacted by adverse childhood experiences. It is built upon affect regulation theory, whereby trained therapists guide youth to explore the influence of their thoughts, feelings, and physiological reactions on sexual decision making. The intervention is composed of ten structured, individual therapy sessions intended to promote optimal sexual decision-making. By increasing knowledge of sexual health and the impact of trauma on sexual decision-making, youth are encouraged and supported in practicing self-regulation.

PS-R uses motivational interviewing as the foundational philosophical approach for the intervention. This involves addressing youth's motivation for change to decrease problem sexual behavior, manage impulsivity, negotiate, and practice harm-free and protected sex. The intervention consists of the TOP Workbook for Sexual Health, sexual education, skill-building, and multi-sensory activities that enhance understanding of key concepts and promote memory retention. PS-R provides a setting for youth and therapists to address the impact of trauma, and how values, beliefs, choices, and personal goals affect a person's sexual health and well-being. It addresses many underlying issues that put youth who have experienced trauma at increased risk for teen pregnancy, STIs, HIV, and sexual harm.

There are three-prongs that make up the intervention: 1) completing the TOP workbook; 2) obtaining sexual health education; and 3) participating in multi-sensory activities to enhance understanding of key concepts, practice skill building, and promote memory retention. Each of the ten structured therapy sessions incorporates these elements in a uniform way to enhance predictability and expectation for change. After the introductory session, the following nine sessions involve: 1) checking-in and facilitating a decision dialogue; 2) monitoring progress in the TOP workbook; 3) providing sexual health education and engaging in multi-sensory activities; and 3) obtaining client feedback.

ii. **Intended program dosage:** *Practice Self-Regulation* is intended to be delivered in 10 one-hour sessions on a weekly basis.

iii. **Intended program content:** PS-R is a theory-guided, therapeutic intervention designed to increase condom use and decrease the number of sexual partners in 14-19 year old adolescents who have been affected by trauma. The content of the intervention reflects the seven essential topics outlined in the National Sexuality Education Standards (anatomy and physiology; puberty and adolescent development; identity; pregnancy and

[2] Schore, A. (2003). Affect regulation and the repair of the self. New York: W.W. Norton & Company.
[3] Detailed information about the PS-R curriculum and theoretical framework can be found in program-related documents, including the *Trauma Informed Approach for Adolescent Sexual Health Conceptual Model* and Schladale, J. (2013). Trauma Informed Approach for Adolescent Sexual Health. Freeport, Maine: (n.p.)

reproduction; sexually transmitted infections and human immunodeficiency virus (HIV); healthy relationships; and personal safety). The workbook contains self-directed activities that help simplify the complex concepts associated with trauma to help youth easily understand and apply effective coping strategies for self-regulation and optimal sexual decision-making. In addition to providing these core components of sexual health education, this intervention provides a setting for discussion between the youth and therapist about the effect of trauma, personal goals, beliefs, values, and choices that affect a person's sexual health and well-being. The intervention also uses components of expressive therapy, such as multisensory activities, to enhance the therapeutic experience for the youth and enable them to envision pathways toward reducing personal risk.

iv. **Intended program delivery:** The intervention is intended to be delivered in a private space (for example, a therapist's office) by master's-level therapists who have been trained on the PS-R program. Ideally, it is delivered over 10 consecutive weeks (one session each week); however, it may take longer depending on a youth's mental health needs and scheduling (for example, vacations or missed appointments).

v. **Program modifications for study purposes:** For the purposes of the study, two modifications have been made to how the intervention is delivered. First, the intervention window has been defined as an 18-week period. During the intervention window, intervention participants are expected to receive all 10 sessions of PS-R in place of their regularly scheduled counseling sessions with their therapist. However, at the therapists' discretion, the PS-R sessions can temporarily halt to address pressing or acute needs of the study participants. Second, to participate in the study, therapists must be: (1) therapists with a master's degree and an active license as a mental health professional; (2) trained in study procedures by a PRG research analyst; and (3) participate in a 24-hour, 3-day PS-R training, conducted by Joann Schladale or another approved PS-R trainer.

b. **Counterfactual condition:** The comparison condition will be business as usual; that is, the typical therapy or counseling that participants would receive from their therapists. There will be no alternative program or additional activities offered to the participants assigned to the comparison group. The comparison condition is referred to as Therapy Practice Group (TPG) for the purposes of the study.

Therapy models typically used by therapists will vary, but it is likely that other trauma-informed interventions or therapies will be used with control participants (e.g., Trauma-Focused Cognitive Behavioral Therapy). Therapists have been instructed they can answer specific questions related to sexual health in therapy sessions; however, they are to provide information only and referrals for additional services when necessary, as they would under typical circumstances. Partner agency representatives and participating private practitioners have confirmed that no other curriculum-based teen pregnancy prevention programs or similar therapy-based sexual education programs will occur during the study period.

The control condition is similar to the treatment in terms of it being a multi-session therapeutic intervention delivered by trained therapists. This conformity with the treatment should reduce any confounds that might arise if the counterfactual experience was different enough in aspects aside from the intended informative and motivational treatment.

i. **Intended program components**: The program components of the control condition will vary depending on the type of therapy provided to the participant, but will not include any sexual education, use of the TOP workbook, or use of multi-sensory activities.

ii. **Intended program dosage**: The control condition is intended to be delivered over the course of multiple therapeutic sessions. The specific number of sessions delivered depends upon the needs of the participant, as assessed by the therapist.

iii. **Intended program content**: The control condition content will vary depending on the type of therapy delivered to the participant, but will not include specific sexual health education.

iv. **Intended program delivery**: The control condition is intended to be delivered in a private space (for example, a therapist's office) over the course of several weeks. All therapists participating in the study are therapists with advanced degrees (for example, a master of social work) and an active clinical license.

3) Study Design

a. **Sample formation:** To become enrolled in the THS sample, adolescents must be receiving therapeutic services from a therapist participating in the study who is trained in the PS-R intervention and on study procedures.

Potential study participants are identified through a recruitment process. Trained therapists working at participating agencies who are familiar with study eligibility criteria first review client databases and complete a *Recruitment and Eligibility Screening Form* for all youth between the ages of 14-19 who are receiving outpatient mental health services from a therapist who is participating in the study. They then call or meet with the potential youth participant to see if the youth is interested in participating and being screened for the study. If the youth is interested, the therapist will then determine whether the individual is appropriate for the study with regards to physical and mental health, whether the individual has ever used the TOP workbook in therapy before, and whether the individual is currently living with any other youth who has enrolled in the study.

To enroll a participant, a PRG study coordinator must conduct a full eligibility screening, which includes obtaining informed consent/assent. If the individual is eligible, they complete the baseline questionnaire. Individuals who take the baseline questionnaire are randomized into the treatment or control condition and considered enrolled in the study**.** This set of participants, who are randomized into the study and offered the opportunity to receive PS-R, constitutes the full intent-to-treat (ITT) sample. The offer to receive the PS-R intervention is the ITT treatment that we investigate in the primary and exploratory research questions.

i. **Eligibility criteria for target population:** a number of criteria have been established for participation in the study.

To be eligible, youth must:

1. Consent/assent to participate;[4]
2. Be at least 14, but not more than 19 years old at the time of their enrollment, as determined by the agency staff or therapist;
3. Be receiving outpatient mental health services at one of the study's implementation sites;

[4] If the youth are 14-17, the youth is considered to have provided "informed assent", as they cannot legally provide written consent. If the youth are 18 or 19, they can provide "informed consent." In Michigan the youth's parent or legal guardian must also be present to go through the informed consent process and sign the form.

4. Be deemed appropriate for the study with regards to physical and mental health by an agency staff member or therapist;[5]

They must not:

1. Be currently enrolled or have previously participated in the Teen Health Study;
2. Have previously used the TOP workbook in therapy;
3. Be living with a youth who is enrolled in the Teen Health Study;[6]
4. Have participated in any other TPP-funded programs;[7]

ii. **Purposeful Sampling:** Any individual who meets initial inclusion and exclusion eligibility criteria, which includes providing informed consent/assent, is asked if they would like to enroll into the study. If they say yes, the study coordinator administers the baseline questionnaire to the study participant and randomizes them into a condition. They are enrolled in the study when they take the baseline questionnaire and are then considered part of the ITT study sample.

b. **Random assignment process**

i. **Unit of randomization:** Random assignment occurs at the individual participant level.

ii. **Random assignment procedure:** Random assignment is conducted during the administration of the baseline questionnaire. Random assignment blocks of varying sizes assign participants to the treatment or control condition at an equal (i.e., 1:1) assignment ratio. Prior to the start of the study in each state, and on an ongoing basis as resupply is needed, PRG prepares 'assignment envelopes' specific to each of the five states, and stratified by gender (male or female), using an existing algorithm available in Stata (random allocation command, ralloc). The allocation lists are produced by a senior analyst, password-protected, and stored on a secure PRG server. Study coordinators in each state are given envelopes (in ascending numerical order) in a box that they assign in sequential order to eligible individuals. Each sealed security envelope has the state name, gender, and unique study ID recorded on the outside and contains a piece of paper indicating the assignment condition (*PS-R* or *TPG*) from the allocation list.

During each baseline questionnaire administration session, the study coordinator picks up the next envelope in the stack (based on the gender of the participant); the number written on the outside of the envelope will be the assigned ID number for that participant. The study coordinator enters this number into the 'study ID' field on the questionnaire (administered on a computer) so that this number is associated with the participant's questionnaire data. While the participant is taking the baseline questionnaire, the study coordinator opens the envelope and reads the paper inside that indicates the condition to which the participant is assigned. This is when random assignment occurs – when the

[5] Youth are deemed inappropriate for the study if the therapist determines that: 1) the youth is psychiatrically unstable; 2) the youth is not cognitively capable of internalizing and understanding the content of the intervention; 3) the youth has engaged in sexual harm towards others, unless the sexual behavior has been successfully treated and the presenting problem does not involve the participant as a sexual offense; 4) the youth does not have enough projected therapy sessions left in order to finish the PS-R intervention; or 5) based on clinical assessment, it is determined that participation may be detrimental to the well-being of the youth.

[6] The one exception to this exclusion criteria is if the two individuals living together are enrolled at the same time into the study. In this scenario, both individuals are treated as if they are being enrolled into the study and go through all screening and enrollment procedures. However, only one person (randomly selected by the research team) is considered a true "study participant." The randomly selected participant is randomized into the study and the condition to which this study participant is assigned is also the condition given to the other person by the therapist.

[7] During the eligibility screening process, youth participants living in certain states are asked whether they have been enrolled into either the TEMPO study (New Mexico only); TPP Planned Parenthood study (California only); Believe in Youth Louisiana Program (Louisiana only); or Youth Empowerment Study (New Mexico only). Individuals who indicate that they have been enrolled into one of these studies are not eligible to enroll into the Teen Health Study.

envelope is opened associating that individual (and ID number) with a treatment condition. As above, the blocking procedure (various sized blocks) will ensure an (almost) equal number of treatment and control assignments in each state. The study coordinator enters the study ID and condition assignment in the *Enrollment Form* while the participant completes the questionnaire.[8]

iii. **Blocking procedures:** Blocking occurs at the state and gender level. Participants are enrolled and randomized at the individual level within each state based upon their self-reported gender.[9]

iv. **Probability of assignment to treatment group**: The probability of assignment to the treatment group is intended to be equal to the probability of assignment to the control condition; that is, p (assignment to treatment) =.5.

v. **Potential for crossover/contamination:** To mitigate potential for both crossover and contamination, the research team works closely with each therapist involved in the study to ensure that the study procedures and expectations regarding randomization are clear before implementation. Therapists commit to ensuring that participants randomly assigned to participate in *PS-R* receive the intervention to fidelity, and participants randomly assigned to control receive therapy as usual. However, despite these efforts, there still remains the possibility that *PS-R* participants may not receive some or all of the intervention sessions, and control participants may receive some *PS-R* content if study procedures are not followed by therapists or if sexual or reproductive health information similar to *PS-R* content is inadvertently delivered to a control participant.

c. **Consent/assent process:** There is no difference in the consent/assent process for the treatment or control groups. Evaluation consent/assent is a condition of eligibility for the study, so no individual is randomized to a condition until after informed consent is obtained.

   After the individual has been screened for initial inclusion and exclusion eligibility criteria, the study coordinator goes through the *Participant Informed Consent Form* with them. This provides the individual with information about the study, outlines why they have been invited to participate, and addresses any questions that may arise. This process involves a paragraph-by-paragraph exploration of the consent form by the study coordinator and the youth and constant "checking in" with the individual to be sure they fully understand the study requirements. At the end of this exercise, they are asked to provide consent/assent if they wish to participate in the study.[10]

d. **Data collection:** Data used for investigating both primary and exploratory research questions are obtained from the *Participant Questionnaire* administered at baseline, post-program, and nine months post-intervention period. The questionnaire is used to collect data on study participants' self-reported contraceptive use and knowledge, sexual behavior and experiences, and intentions, thoughts, and feelings related to sexual behaviors. It is administered three times at the following time points:

[8] Each ID number and its corresponding intervention assignment is logged by the study coordinator on the participant's *Enrollment Form*. ID numbers and assignments from the *Enrollment Form* dataset are then matched to PRG's randomization allocation dataset so that we can monitor the integrity of the randomization process. This should ensure, at a minimum, that the condition a particular participant is assigned is the one that is indicated in the assignment records. This is to say that the ITT "point of offer" treatment should, at minimum, formally retain all the properties of random assignment even if a therapist wrongly administers the incorrect intervention.

[9] Any client who identifies themselves initially by a gender other than male or female is asked whether they have strong preference towards male or female. If they remain gender neutral, then the study coordinator asks them for their biological sex and classifies them in this way for the purposes of our study design.

[10] See footnote 4.

a. Baseline – at the enrollment session just prior to the participant receiving their assigned intervention
b. Post-program – following the close of the intervention-period (18 weeks after the baseline questionnaire administration)
c. Nine months post-intervention period – nine months after the 18-week intervention period (or 13.5 months after the baseline questionnaire administration)

While we collect data at three times over a period of 18 months, our analysis of primary research questions is concerned only with data gathered at baseline and nine months post-intervention.[11] Exploratory research questions will investigate mediating factors, subgroup effects, and other, exploratory outcomes. Study participants who are offered *PS-R* and *TPG* receive the same questionnaire. The questionnaire contains 115 items and takes, on average, 18 minutes to complete. The instrument was constructed by PRG staff and is composed of items and scales that have been used in previous research on sexual behaviors and contraceptive use. The instrument was reviewed by health professionals and pilot-tested by youth with similar characteristics to our proposed study population. The questionnaire includes the same items at each time point and will measure the same constructs with identical measures at each administration.

There is only one difference in the data collection procedures for treatment and control groups. At the post-program time point, treatment participants take a feedback questionnaire after they have completed the *Participant Questionnaire.* The feedback questionnaire asks treatment participants to report on their experiences with the *PS-R* intervention. Otherwise, data collection is conducted identically for both groups.

There are minor variations in procedures at the different data collection points. For all data collection points, we administer the questionnaire using a web-based (online) survey administration tool that has Audio Computer Assisted Self-Interview (ACASI) capabilities. The baseline and follow-up online questionnaires are administered by PRG study coordinators. Study coordinators make every attempt to collect outcome data as soon as possible after each data collection window (post-program and nine-month follow-up) opens; however, the data collection window remains open for four months to allow sufficient time for participants to complete their questionnaires. Any questionnaires completed after a data collection window closes will not be included in the final analytic sample.

At each data collection point, the study coordinator identifies a quiet and private space for the participant to complete the questionnaire. Questionnaires are completed in locations that ensure participant confidentiality as well as convenience, such as the therapist's office space, coffee shops, libraries, or other public locations like university buildings. On the computer, the study coordinator selects the appropriate link to the type of questionnaire to be administered, enters the participant's unique participant ID number, and gives the study participant some brief instructions about taking the questionnaire. The script for the instructions was developed by PRG and emphasizes the importance of the participant's honesty in answering questions and the confidentiality of their responses. Study coordinators are instructed to read it prior to each questionnaire administration. The study coordinator also provides the participant with a calendar that they can reference when taking the questionnaire.

[11] With the assumption that we maintain low attrition and that the RCT is executed with integrity, we could approximate an un-biased estimate of the average treatment effect of *PS-R* by comparing differences in the means of our outcome variables reported by the treatment group with those reported by the control group. We could then provide a compelling response to our research question by testing the hypothesis that there is no difference between the two groups using straight-forward hypothesis testing statistics (t-test). However, we propose to use regression-adjusted means as the primary estimate of *PS-R* program effects to improve the precision of our estimates. Refer to subsection 4)ef below for a more detailed description of our proposed analytic approach.

The study coordinator then tells the participant to click "Next" on the computer screen when ready to begin and to click on the "Submit" button when they have finished the questionnaire. They then turn the study computer over to the participant to complete the questionnaire. The study coordinator may be present in the room while the participant completes the questionnaire, but study staff are trained to minimize the impact of their presence and to clearly explain that the reason they are staying is to facilitate questionnaire administration.[12] After each questionnaire administration, the study coordinator receives an automatic email letting her know the questionnaire has been completed.

If, for some reason, the internet is not working when a questionnaire needs to be completed, or if the participant's needs dictate that she or he takes the questionnaire on paper (e.g., due to computer illiteracy), the study coordinator follows the same procedures as above, but the study participant completes the questionnaire using a self-administered paper version. The study coordinator enters the study participant's unique ID on the paper questionnaire and gives them the questionnaire and a large envelope. At the end of the regular questionnaire instructions, the study coordinator instructs the participant to put their completed questionnaire into the envelope and seal it when finished. The participant then hands it to the study coordinator who writes the study ID number on the outside of the envelope and stores it in a locked file cabinet. Paper questionnaires are mailed via FedEx to PRG on a regular basis and entered by PRG staff into the *Participant Database*.

Research staff attempt to administer all questionnaires in person; however, two weeks after the data collection window opens, if a participant is unwilling or unable to complete follow-up questionnaires in person, the participant is given the option to complete it online on a personal device (computer or tablet) using a survey link provided via email. Two months after a follow-up window has opened, if study coordinators have not been successful in getting a participant to complete a questionnaire, they are allowed to then offer a $60 incentive (an increase of $20 from the original $40 incentive). Three months after a follow-up window has opened, if study coordinators have not been successful in getting a participant to complete the questionnaire online or in-person with a higher incentive offer, the final option offered is to complete a shorter version of the questionnaire over the phone, in an interview format, with the participant. We will run sensitivity analyses that exclude participants who were surveyed by phone from our analytic sample and report substantive differences in the results section of the report.

e. **Data collection related to additional analyses:**

Recruitment Log

The *Recruitment Log* is an electronic database housed in the *Zoho Creator* application. Data from the paper *Recruitment and Eligibility Screening Form* are entered into the electronic *Recruitment Log* by the study coordinator. Each youth who is screened for the study is entered as one record in the log. The *Recruitment and Eligibility Screening Form* is a paper form used by agency staff and therapists to collect eligibility information in person from potential study participants who come into the clinic for therapy. The form includes a study introduction script, eligibility screening questions to be asked of the individual, and collects eligibility determination data. Data from these forms are entered into the *Recruitment Log* database by study coordinators on an ongoing basis.

Participant Database

The *Participant Database* is an electronic database housed in a *Zoho Creator* application. Data from the paper *Enrollment Form, Post-Program Data Collection Form,* and *9-Month Follow-up*

[12] There may be some occasions when the study coordinator needs to briefly leave the room, but in these instances, the study coordinators have been instructed to still remain available to participants.

Data Collection Form are entered into the electronic *Participant Database* by study coordinators. Each participant enrolled in the study should have one record in this database for each of the three forms housed here (*Enrollment Form, Post-Program Data Collection Form,* and *9-Month Follow-up Data Collection Form*). The paper *Enrollment Form* collects administrative data, participant information on language preference, data on adherence to the study procedures, questionnaire completion data, baseline incentive tracking numbers, and any notes on issues/changes to the study protocol. The paper *Post-Program* and *9-Month Follow-Up Data Collection Form* forms are completed by the study coordinator at the post-program and 9-month follow-up data collection points. One form should be completed for each enrolled participant. The forms collect administrative participant information, questionnaire completion data, incentive tracking data, and notes on issues/concerns with the post-program or 9-month questionnaire administration session.

Attendance Form

This is a paper form created by PRG that is completed by therapists during the intervention phase of the study. One form is completed for each study participant. For participants assigned to receive the PS-R intervention, therapists record the session date, whether PS-R was provided, the number of the PS-R session plan that was provided (if applicable), reason why PS-R was not provided (if applicable), duration of the session, whether the session was recorded, and any referrals provided during the session on the form each time they meet with a study participant for an individual outpatient therapy session. For participants assigned to receive the control condition, therapists record the session date, duration of the session, whether the session was recorded, and any referrals provided during the session each time they meet with a study participant for an individual outpatient therapy session. Data from this form are entered into the *Attendance Log* by study coordinators on an ongoing basis and stored on the PRG server. The paper forms are stored in a locked filing cabinet.

Therapist Self-Report Form – PS-R and Control Group

These two paper forms created by PRG are completed by therapists during the intervention phase of the study. The *PS-R Therapist Self-Report Form* is completed when a PS-R session is delivered to a participant, and the control group form is completed when a control therapy session is delivered to a participant. The PS-R form tracks administrative data, the number of activities completed at each session, any adaptations to activities, why activities were not completed, and notes on issues with the implementation of the intervention session. The control group form tracks whether any PS-R content was delivered and the type of therapy provided during the session. Data from these forms are entered into the *Fidelity Monitoring Database* by study coordinators on an ongoing basis and stored on the PRG server. The paper forms are stored in a locked filing cabinet.

4) Analysis

a. **Outcome measures:** Our primary research questions ask to what extent the offer to participate in *PS-R* relative to the offer to participate in *TPG* impacts participants' reported: 1) times having sex without a condom; and 2) number of sexual partners nine months after receiving the intervention. We describe below the specific operationalization of these two outcome measures.

Times having sex without a condom

We operationalize times having sex without a condom as a risk outcome; that is, we measure the frequency with which participants engage in the risk behavior of having sex without a condom, rather than the frequency with which they engage in the safe sex practice of using condoms.

Constructing the variable in this way allows us to examine the self-reported sexual behaviors of the full analytic sample of participants, regardless as to whether or not they are sexually active.

Specifically, times having sex without a condom is constructed as a continuous variable – the number of times in the past three months a participant *does not* use condoms while engaging in any type of sex.[13] Data used to assess the impact of the treatment (*PS-R*) on condom use are obtained from the following three items on the *Participant Questionnaire,* which is administered to both the treatment and control groups at baseline and nine months post-intervention.

- *In the past three months, how many times have you had vaginal sex without using a condom?*
- *In the past three months, how many times have you had oral sex without using a condom?*
- *In the past three months, how many times have you had anal sex without using a condom?*

Persons who indicate that they have not ever had a particular type of sex (vaginal, oral, or anal) or have not had that type of sex in the past three months are coded as having that type of sex without a condom zero times. [14] The final outcome measure is calculated by summing individuals' responses to these three items (the number of times they report they did not use condoms during vaginal, oral, and anal sex).

PS-R will be considered to have a positive impact on times having sex without a condom if the number of times having sex without condoms reported by participants assigned to *PS-R* at the nine-month follow-up is smaller than the times reported by control participants and the difference between groups is statically significant.

Number of Sexual Partners

Number of sexual partners is constructed as a continuous variable – the number of sexual partners the participant reports that they have had in the past three months. Data used to assess the impact of the treatment (*PS-R*) on number of sexual partners are obtained from the following two items on the *Participan*t *Questionnaire,* which is administered to both the treatment and control groups at baseline and nine months post-intervention period.

- The first question asks: *During your life, with whom have you had sexual contact?*
- If respondents select *Females, Males,* or *Females and males* to the first question, they are then asked this second question: *How many sexual partners have you had in the past 3 months?*[15]

Persons who indicate that they have never had sexual contact or have not had any sexual partners in the past three months are coded as having zero sexual partners. [16]

PS-R will be considered to have a positive impact on number of sexual partners in the past three months if the number of sexual partners reported by participants assigned to *PS-R* at the nine-

[13] *PS-R* is aimed at reducing STI risk through the development of condom and STI knowledge. Since STIs can be transmitted through any type of sexual contact (i.e., vaginal, anal, or oral), our measure of times having sex without condom use is not limited to sexual intercourse but includes all self-reported sexual activity.

[14] The *Participant Questionnaire* contains sexual behavior questions that use a three-month recall period. As research has consistently found that memory of behaviors/events decreases over time and accuracy of recall is negatively associated with length of recall period (Clarke et al. 2008; Schwarz and Oyserman 2001), we use items with three-month recall periods to construct our measures of sexual behaviors since these should elicit more accurate responses than a longer recall period (e.g., six-month).

[15] The alternative response to this first question is *I have never had sexual contact.* If a participant selects this response, they are skipped out of the subsequent question, *How many sexual partners have you had in the past 3 months*?

[16] See footnote **Error! Bookmark not defined.**.

month follow-up is smaller than the number of sexual partners reported by control participants and the difference between groups is statically significant.

b. **Analytic sample:** In California, Louisiana, Maine, Michigan, and New Mexico, 79 therapists working in eight mental health agencies and 89 independent therapists have been recruited to identify youth who are eligible to participate in the study. As explained in the *Study Design* section, agency staff and private practitioners review administrative/clinic data for patients receiving therapy services to identify potentially eligible participants.

   Potential participants are eligible if they: 1) are between the ages of 14 and 19 years of age; 2) are receiving outpatient mental health services at one of the participating study sites; 3) consent/assent to participate; 4) are not currently enrolled or have previously participated in the THS; 5) have not previously used the TOP Workbook in therapy; 6) are not living with a youth who is enrolled in the THS;[17] 7) have not participated in any other TPP-funded programs; and 8) are deemed appropriate for the study with regards to physical and mental health by an agency staff member or private practitioner.

   After eligibility is determined, participants are randomized into the treatment (*PS-R*) or control (*TPG*) condition according to procedures detailed in subsection 3b. The act of randomization into either the treatment or control arm constitutes the offer to participate in the intervention and is the point at which the individual becomes a participant in the study. The analytic sample is defined as all participants who were randomized into either the treatment or control conditions and who have reported sufficient outcome and covariate data.[18] Missing data procedures are outlined in subsection 4c*v* below.

c. **Data cleaning:** Prior to analysis, PRG staff will systematically screen or review the analytic variables (baseline and outcome) to identify invalid, inconsistent, outlying, missing, and unreliable data.[19] New variables are created in which data that are deemed unusable (i.e., invalid or unreliable) are coded as missing and flagged according to missing data type; all other data are retained, unchanged.[20] The steps taken in this data cleaning process are outlined below.

   i. **Identify and flag unreliable cases**: The first step in the data screening process is to identify and flag entire cases (i.e., entire questionnaires) that are unreliable. By unreliable, we mean that we have sufficient reason to believe that the respondent's answers are not honest representations of their behaviors, knowledge, and beliefs. These cases are treated as missing and excluded from our benchmark analyses.

      Cases are flagged as unreliable when responses follow a clear, deliberate pattern. This data cleaning procedure is informed by the data processing rules established for the National Survey on Drug Use and Drug Health (NSDUDH) and for the Youth Risk

[17] See footnote 6.
[18] As outlined in subsection 4*cv*, our benchmark approach is to impute baseline/covariate data. As such, sufficient baseline/covariate data means all cases where data are not unit missing. We do not anticipate that we will have different analytic samples for our outcomes of interest; data are expected to be missing entirely for any given respondent at any observation point or not. If for whatever reason analytic samples are different for different outcomes, we will assess baseline equivalence separately for each analytic sample.
[19] We propose to document the prevalence of inconsistent and missing data in a descriptives table presented as an appendix in our final impact report. Along with our presentation of sensitivity analyses, we will present tables that present the prevalence of unit and item missing (which result from nonresponse) as well as inconsistent, unreliable, and invalid data for both treatment and control samples. Regarding inconsistencies specifically, for each sexual behavior variable included in our model specifications (which could therefore influence the constitution of the analytic sample) we will include the following: sample size (the number of observations prior to recoding of inconsistencies) and the number of observations that are inconsistent over-time. If paper questionnaires lead to internally inconsistent data, we will also report on this.
[20] A note on missing values: Stata provides a series of missing value codes that allow us to "flag" missing data according to why they are missing. Data that are missing due to unit nonresponse (a questionnaire was not completed) are coded using the "system missing" value (".") . All other types of missing data are coded using "extended missing" values (e.g.,".a", ".b").

Behavior Survey (YRBS), which treat records that follow defined patterns of responses as missing.[21] PRG flags the following cases as unreliable: a) the same response option is chosen for all multiple choice questions; b) responses alternate between only two response options; or c) responses alternate systematically, starting with one response option, alternating through all options in order until exhausted then beginning again (in the same or in reverse order). If other response patterns are observed over the course of the evaluation, they will be added to PRG's list of unreliable response patterns.

Data for cases that are deemed unreliable are treated as *unit missing* and excluded from benchmark analyses. However, sensitivity analyses that include the unreliable data will be conducted and results will be reported in an appendix of the report.

ii. **Identify and flag invalid responses**: The second step in the data screening process is to inspect the data for instances in which responses are invalid because they are outside of a pre-determined range of plausible or acceptable values. Each questionnaire type (e.g., baseline, post-program) has a codebook, which is prepared by a PRG staff, that contains variable names, valid variable values or ranges of values, and when applicable value labels.[22] Referring to the codebook, a senior or lead research analyst performs diagnostics in Stata to ensure that responses to all analytic measures are valid (i.e., data are within ranges specified in the codebook). A data analyst inspects the data using two commands in Stata. First the analyst uses the command *sum variable_name,* which provides summary statistics (mean, minimum, maximum, standard deviation) for all numeric variables. The analyst checks that the minimum and maximum values are valid. If this command reveals there are values out of range, the analyst then inspects the data using the command, *tab variable_name, missing*, which provides a frequency table of all values (including missing values) so the analyst can identify and flag all values that are out of range as invalid and recode these values to missing (code as ".i").

Data that are recoded to missing are treated according to our missing data approach. Briefly, our benchmark approach is to impute or adjust missing baseline data and include in analysis; we exclude observations with missing outcome data from analysis.

iii. **Identify and flag outliers:** The third step is to identify and flag severe outliers. Outliers (operationally defined below) are values that are extreme compared to other observations but are not plainly invalid. In the data cleaning process, we inspect outliers so that we can try to ascertain whether they are in fact true (or plausible) values or potentially a result of measurement error. The only variables for which we inspect outliers are those used in the construction of our outcome variables (times having vaginal sex, times having vaginal sex without a condom, times having oral sex, times having oral sex without a condom, times having anal sex, times having anal sex without a condom, and number of sexual partners) because they have no upper limit (all other variables used in analysis are either categorical or have predicated upper and lower bounds). Our approach to identifying and flagging outliers is as follows.[23]

- First, in Stata we use the *lv* (letter-value display) command to identify severe outliers. We define values as severe outliers according their relation to the

[21] See Comparing and Evaluating Youth Substance Use Estimates from the NSDUH and Other Surveys retrieved February 27, 2018 from https://www.samhsa.gov/data/sites/default/files/NSDUH-M9-Youth-2012/NSDUH-M9-Youth-2012.pdf.
[22]Regardless as to whether data are nominal, ordinal, or continuous, all response options are coded in Stata as numeric values; values are labeled according to corresponding category names when data are nominal or ordinal. As an example, the variable gender is a nominal variable; however, it is treated as a dummy variable where females are coded as "1" and males are coded as "0". The only acceptable values for this variable then are 0 and 1; any other values are out of range.
[23] Rules for identifying outliers are informed by the following: Hamilton, Lawrence C. 2006. *Statistics with Stata: Updated for Version 9* and *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/.

interquartile range (IQR). Severe outliers are defined as values outside of the *outer fences* of the population distribution.
    - IQR= Q3(3rd quartile or 75th percentile) – Q1(1st quartile or 25th percentile)
    - Upper outer fence: Q3 + 3*IQ
    - Lower outer fence: Q1 - 3*IQ
- Second, we create an outlier indicator variable, where observations deemed severe outliers are coded as 1, all others are coded as 0.

Our benchmark analytic approach is to include data flagged as outliers in analysis, because we do not know for certain whether the values are true or invalid. However, we also run sensitivity analyses that exclude these data and report substantive differences in the results section of the report.

iv. **Identify and flag inconsistencies in reporting of sexual behaviors**: The fourth step in the data review process is to inspect the data and identify inconsistencies in sexual behavior outcome data. With repeated measures of sexual behaviors, two primary types of inconsistencies may occur – internal inconsistences and over-time inconsistencies.[24] Internal inconsistencies refer to discrepancies in responses (to related questions) in the same survey administration. For instance, a respondent might say that they have not had sex in the past three months, but then indicates that they used condoms three of the times they had sex in the past three months. Over-time inconsistencies refer to instances in which lifetime reported behaviors decline or are completely recanted over time. For example, at baseline a respondent might say that they have had vaginal sex in their life, but on a subsequent administration of the survey say that they have never had vaginal sex.

In order to minimize internal inconsistencies in our primary outcomes, we built skip patterns into the online questionnaire – if participants indicate they have not had a particular type of sex they are skipped out of more specific questions related to that type of sex; and if they state they have never had sexual contact, they are then skipped out of questions asking about how many sexual partners they have had over certain periods of time. In addition, participants are precluded from indicating they had a particular type of sex without a condom more times than they said they had that type of sex. Because of this, no internal inconsistencies can exist in our primary outcome measures.[25]

To address over-time inconstancies, a research analyst examines all variables that are used to construct primary outcome measures, as well as any variables that may be used to logically impute values for primary outcome measures. If over-time inconsistencies are identified, both the baseline and follow-up values are flagged as inconsistent over time and recoded to missing (coded as ".k."). Data that are recoded to missing are treated according to our missing data approach. Briefly, our benchmark approach is to adjust

[24] Inconsistencies can occur for a number of reasons including social desirability bias and memory or recall issues on the part of the respondent and misunderstanding on the part of either the respondent or interviewer (Alexander et al 1993; Clarke, Fiebig, and Gerdtham 2008; Del Boca and Noll 2000; Harris et al 2008; Schroeder et al 2003; Schwarz and Oyserman 2001). These issues are especially common in self-reports of sexual behaviors where questions are of a sensitive nature and often respondents are asked to indicate the frequency and/or recency of behaviors over differing lengths of time (e.g., 30 days, 3 months, 6 months).

[25] In rare instances, paper-based questionnaires are completed by participants. This may occur if the study coordinator is unable to open the online questionnaire because of network connection issues, or if the participant cannot be reached for an online administration and is instead mailed a paper questionnaire. Paper-based questionnaires include the same skip patterns as online questionnaires; however, unlike the online questionnaire where participants are skipped out of subsequent questions based on their response to previous questions and cannot report inconsistent times having different types of sex or number of sexual partners, it is possible for participants to provide responses to questions that they should not have answered or that do not make sense given their responses to previous questions. If and when such inconsistencies are observed in data originating from paper-based questionnaires, the participant's responses to the two or more items where inconsistencies are noted are recoded to system missing and will not be used in our primary outcome analysis.

missing baseline data and include in analysis; we exclude observations with missing outcome data from analysis.

v. **Missing data approach:** Assuming that our study design and procedures are sound, missing data pose perhaps the greatest threat to the internal validity of our RCT study and the ITT framework (Puma et al. 2009; Moher et al., 2010).[26] Randomization at the point of offer allows us to make causal statements about the effect of that offer because treatment and comparison samples are equal in expectation. For the ITT framework to remain internally valid, however, the treatment and comparison groups must remain equal in expectation at the point of analysis. When the analytic sample is diminished by attrition or non-response, non-random differences (i.e., self-selecting) between the treatment and comparison groups may be introduced into the sample and estimates of program impacts may become biased. Although there is no consensus on how to resolve this, practical guidance on how to address and mitigate the problems associated with missing data have been published in education (Puma et al., 2009).

Our six-step decision process for addressing this problem, as detailed below, is informed by this guidance. These steps articulate how we will deal with missing outcome and baseline/covariate data (that is variables outlined in the *Model specification and covariates* section and are necessary for the estimation of impacts). The benchmark approach that we have selected aims to mitigate the introduction of bias into our impact estimates and maximize the use of available data by adjusting missing baseline/covariate data. To test the robustness of this approach, and to verify these findings, we will report comparative findings using sensitivity analyses that also employ an alternative method which includes no adjustment (as outlined in step 6).

1. Using data cleaning procedures outlined in the *Data cleaning* section, identify inconsistent, outlying, unreliable, and invalid data in any analytic (i.e., outcome, baseline, or covariate) variables, recode inconsistent and invalid data as missing, and flag unreliable and outlier data for analysis.[27]

2. Report prevalence of unit and item missingness (which result from nonresponse) as well as inconsistent, unreliable, and invalid data for both treatment and control samples.[28]

3. Determine if logical imputations are possible for any analytic variables that may have missing values (due to nonresponse) and logically impute where this is the case. We will not logically impute where the missing values are previously inconsistent, unreliable, or invalid.

4. Determine if any individuals who are in the randomized sample (for each outcome) do not have outcome data at the nine-month follow-up time point. If this is the case, our proposed benchmark approach is to use case deletion, as we feel it is the most straightforward and prudent approach for missing follow-up data

[26] Puma, M.J., Olsen, R.B., Bell, S.H., Price, C. (2009). What to Do When Data Are Missing in Group Randomized Controlled Trials. (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Moher, D. et al. (2010). CONONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials. BMJ 2010;340:c869.

[27] We will code missing responses with a unique missing code that identifies or flags these missing values according to the reason they are missing (i.e., nonresponse, invalid, inconsistent). Unreliable data are not recoded to missing, rather cases deemed unreliable are coded as 1 in an indicator variable, treated as unit missing, and excluded from analysis. See the *Data cleaning* section or Table 3 in the Appendix for details on how missing data are coded.

[28] For item missing values, we will only report prevalence of missing and inconsistent data for variables that are included in our model specifications and could therefore influence the constitution of the analytic sample.

recommended in Puma et al. (2009). These cases will be deleted from the analytic sample and attrition statistics will be reported.

5. Determine if any individuals who are in the randomized sample (for each outcome) are missing baseline covariates or the baseline measure of the outcome variable. If this is the case, our proposed benchmark approach is to use dummy variable adjustment procedures, as we feel it is the most straightforward and prudent approach for missing baseline/covariate data recommended in Puma et al. (2009).

6. Conduct sensitivity analyses by estimating results with missing baseline data excluded from the analysis (i.e., use case-wise deletion for all cases with missing baseline and outcome data). In an appendix, we will report our benchmark results next to the sensitivity analysis results to verify findings.

d. **Assessment of baseline equivalence:** Baseline equivalence will be reported for all baseline measures of each outcome variable as well as relevant demographic and sexual behavioral measures. We first list and describe the measures we will use to examine the equivalence of our treatment and control group at baseline. After we identify the measures, we provide details on the diagnostic methods that we will use to assess any baseline difference that may exist between the treatment and control groups in the measures outlined below.

Demographic and Sexual Behavior Measures

Baseline equivalence will be assessed for five demographic variables and one baseline measure of sexual behavior (identified below).[29] Age is constructed from data reported by agency staff or therapists in the *Recruitment and Eligibility Screening Form*. Race, ethnicity, and education are constructed using participant self-responses to questions in the THS baseline *Participant Questionnaire*. For the race variables, categorical responses to a single question are used to create multiple dichotomous variables. The sexual behavioral variable is constructed from participants' responses to three questions in the baseline *Participant Questionnaire.* We provide details on variable coding below; details on variable construction can be found in Table 2 in the Appendix.

Demographic:

- Age at screening (continuous; range 14-19)[30]
- Race[31]
- Hispanic/Latino
- Completed high school (0 = has not completed high school; 1= has completed high school)

Sexual Behavior:

- Ever had sex (0=never had vaginal, anal, or oral sex; 1=has had vaginal, anal, and/or oral sex in lifetime);

[29] Baseline equivalence will also be reported for gender out of convention. However, gender is a blocking variable used in our randomization process, so the proportions of individuals by gender who are assigned to treatment and control will be balanced by design. Gender at screening is determined using the participant's gender, as reported by agency staff or therapists.

[30] Age at screening is determined using the participant's date of birth, as reported by agency staff or therapists.

[31] At baseline, participants are asked "What is your race?" and are provided with a list of the following response options: *White; Black or African American; Hispanic, Latino, or Spanish origin; American Indian or Alaskan Native; Asian; Native Hawaiian/Other Pacific Islander; Unknown;* or *Some other race/ethnicity*. Participants can select more than one category and they can also specify some other race/ethnicity. This item is used to create two separate categorical variables. Hispanic/Latino is a dummy variable that indicates whether someone identifies as *Hispanic, Latino, or of Spanish origin* or not. Race is a categorical variable that indicates a person's self-identified race; responses are recoded into the following mutually exclusive categories: *White only, Black only, Other race only* (which includes *American Indian/Alaskan Native, Asian, Native Hawaiian/Other Pacific Islander*, *Other*), *Race not selected* (individuals who selected no racial category), and *Multiracial* (which includes individuals who selected more than one racial category). For analysis, dummy variables are created for each category in the recoded variable.

Baseline Outcome Measures

In addition to the demographic and sexual behavior measures, we will assess baseline equivalency of the outcome measures. We provide details on variable coding below; details on variable construction can be found in Table 2 in the Appendix.

- *Times having sex without condom in the past 3 months* at baseline (continuous; values range 0 to *k*, where 0= has had sex without condoms 0 times in past 3 months and *k*= number of times having sex without condoms in past 3 months)
- *Number of sexual partners in the past 3 months* at baseline (continuous; values range 0 to *k*, where 0= has 0 sexual partners in past 3 months and *k*= number of sexual partners in past 3 months)

Balance Assessment Methods

We propose to assess baseline equivalence of the treatment and control groups according to a multi-step procedure. Baseline equivalence statistics will be produced for each analytic sample.[32] Only participants who provide baseline data to an outcome measure will be included in the analytic sample for that same outcome measure; thus the analytic sample used for each research question may vary slightly because of the exclusion of non-responders. As required by the "Identifying Programs that Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors" review protocol, we will report the adjusted means and p-values of the differences in the baseline variable of interest for the treatment and control groups.[33] We will also report the standardized mean difference of each baseline variable for the treatment and control groups. This last statistic is not required by the review protocol but it is more consistent with the literature on balance statistics.[34]

To establish baseline equivalence, we propose to generate model-based point estimates of the difference between the treatment and control groups for the identified baseline equivalence variables. We will report the adjusted means and p-values of the differences in the baseline variable of interest for the treatment and control groups. We will then compute the pooled standard deviation of these variables. Finally, we will produce a standardized difference of means by dividing the first term by the second.[35]

> ***step 1.*** First, we generate a model-based estimate of the difference between treatment and comparison groups on the pre-intervention measures identified above. Separate models will be run for each of the variables. The empirical model will be estimated with OLS (using Stata). If the measure is dichotomous, we propose to use a linear probability model to estimate the predicted probability of group membership. The model is a reduced-form variation of the model that we use to estimate program impact (as detailed in the *Model specification and covariates* section, below).[36]
>
> $$Y_{baseline} = \beta_0 + \beta_1 T + \sum(\beta_P X_P) + \varepsilon$$
>
> where:

[32] Due to item missing outcome data, we expect there may be slight differences in analytic samples for each research question.
[33] Goesling, B., & Trenholm, C. (2016). Identifying Programs that Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors. Mathematica Policy Research.
[34] The literature on balance statistics argues that significance testing is inappropriate for this diagnostic task (Austin, 2007; Imai et al., 2008; Austin, 2009; Stuart, 2010). Hypothesis tests can be misleading diagnostic measures of baseline equivalence because they conflate balance with statistical power.
[35] Note that we will produce diagnostic estimates of baseline equivalence on the exact same samples of observations that we will use in our primary analysis. In other words, we will apply the missing data approach outlined in section 4c*v* prior to producing estimates of baseline equivalency on the pre-intervention measures.
[36] It is a reduced- form because individual-level, demographic covariates are omitted. It is a variation because the dependent variable is the baseline equivalence variable, not the outcome measure.

$Y_{baseline}$ – is the baseline measure of the variable that we use to establish baseline equivalency (identified in the Appendix – Table 2 above). This variable is included as a covariate in the analytic model (see Table 2 in the Appendix for details on variable coding). Separate models will be estimated for each baseline equivalency measure specified above.
T – A dummy treatment indicator variable whose value equals 1 if the participant is randomized into the treatment group and zero otherwise.
X – A p vector of blocking variables (i.e., subgroups within which random assignment occurred) to account for the variation in the baseline variables associated with these groups. Blocking variables include:

a) State – An n-1 vector of state indicator dummy variables that are coded one if the intervention was delivered at state *n* and coded zero otherwise.
b) Female – A dummy variable that is coded 1 if the participant self-reported being female at screening and coded 0 otherwise.

β_0 – The intercept term, which represents the adjusted mean value of the baseline equivalency measure for participants in the control sample, with all other variables in the model held constant at zero.
β_1 – This represents the adjusted (but not standardized) mean difference in the baseline equivalency variable between treatment and control participants.
ε – The residual or random variation that remains for each observation after the structural components of the model are estimated. It is the difference between the observed and the predicted values at the individual level.

step 2. Report the adjusted means and p-values of the differences in the baseline variable of interest for the treatment and control groups.

step 3. If the pre-intervention measures is continuous, we propose to use the following formula to calculate the pooled within-group standard deviation of the outcome measure:

$$S_p = \sqrt{\frac{(n_t - 1)S_t^2 + (n_c - 1)S_c^2}{(n_t + n_c - 2)}}$$

where: n_t and n_c are the sample sizes, and S_t and S_c are the participant-level standard deviations for the pre-intervention measures for the analytic treatment and comparison groups, respectively. We will produce separate calculations of the pooled standardized deviation for each variable used to establish baseline equivalence (as noted above).

step 4. Produce the standardized difference of means. If the pre-intervention measure is continuous, we will use Hedges' g as the formula to compute the standardized difference of means for the treatment and comparison groups:

$$g = \frac{\beta_1}{S_p}$$

Where: β_1 is the adjusted mean difference in the variable selected to establish baseline equivalence for the treatment and comparison groups (calculated in Step 1), and S_p is the pooled standard deviation (produced in Step 2).

For dichotomous baseline variables we will use the Cox index, which yields effect size values similar to the values of Hedges' g that one would obtain if group means, standard deviations, and sample sizes were available, assuming the dichotomous outcome

measure is based on any underlying normal distribution." Following this guidance, we propose to use the Cox index to estimate baseline equivalence for dichotomous baseline covariates. The formula is as follows:

$$d_{Cox} = \left[ln\left(\frac{p_t}{1-p_t}\right) - ln\left(\frac{p_c}{1-p_c}\right)\right] \Big/ 1.65$$

Where: pt and pc represent the probability of occurrence of the event (or characteristic) within the treatment and comparison groups, respectively.

e. **Condition crossover and contamination:** Crossover will be defined as study participants assigned to the treatment condition who did not receive any *PS-R* sessions over the course of therapy.[37] This will be determined from attendance records within the *Attendance Form*. We will calculate crossover using the following formula:

$$Crossover_{participant}\, T = \left(\frac{Not\ Received_{participant} T}{Base_{participant} T}\right)$$

where:

$Crossover_{participant} T$ - the proportion of participants randomly assigned to the treatment group who did not receive any PS-R sessions

$Base_{partcipant} T$ - the number of participants randomly assigned to the treatment group

$Not\ Received_{participant} T$ - the number of participants randomly assigned to the treatment group who did not receive any PS-R sessions

Contamination will be defined as study participants assigned to the control condition who received any amount of PS-R content over the course of therapy.[38] This will be determined from therapist records within the *Therapist Self-Report Tool*. We will calculate contamination using the following formula:

$$Contamination_{participant}\, C = \left(\frac{Received_{participant} C}{Base_{participant} C}\right)$$

where:

$Contamination_{participant} T$ - the proportion of participants randomly assigned to the control group who received any amount of PS-R

$Base_{participant} T$ - the number of participants randomly assigned to the control group

$Received_{participant} T$ - the number of participants randomly assigned to the control group who received any amount of PS-R

Levels of crossover and contamination will be reported in the findings section of our final impact report.

[37] In the OAH Impact Analysis Plan guidance for Cohort 2 Tier 2B grantees, crossover is described as occurring "*when individuals randomly assigned to the intervention or counterfactual conditions are later found to be receiving the services intended to be offered to the other condition*." Given this, we calculate crossover in our sample as participants assigned to the treatment condition who only received business-as-usual therapy sessions and did not receive any PS-R sessions, as this is the intended counterfactual condition.

[38] In the OAH Impact Analysis Plan guidance for Cohort 2 Tier 2B grantees, contamination is described as occurring "*when individuals assigned to the counterfactual condition end up receiving all or portions of the conditions intended only as part of the intervention*." Given this, we calculate contamination in our sample as participants assigned to the control condition who received any amount of PS-R content, as this is the intended treatment condition.

f. **Analytic approach for primary research questions:** As detailed in our primary research questions, this study investigates whether offering *PS-R* to participants impacts their reported number of times having sex without condoms and number of sexual partners. We do this within the intent to treat (ITT) framework, which does not measure the effect of the participant's exposure to the treatment itself but rather the effect of the offer of the treatment relative to the offer of receiving the control condition. This framework maintains the integrity of the experimental structure by including all participants who were randomized (except those who attrite) in the analytic sample, thereby maintaining an exogenous assignment of participants to experimental condition. Bias can be insinuated, however, through self-selection if any participant who is randomized fails to provide outcome data.

i. **Model specification and covariates:** The primary research questions under investigation in this study are whether offering *PS-R* to participants impacts their: (1) reported times having sex without condoms, and (2) reported number of sexual partners (see Table 1 in Appendix for variable constructions). We propose to estimate these impacts using a regression that will model intervention effects as a function of assignment to *PS-R* (i.e., Treatment), relevant baseline covariates, a baseline measure of the outcome variable, and gender and state-level (blocking) indicators (see Table 2 in Appendix for variable constructions). Although a straight difference-of-means approach should provide unbiased estimates of the effect of the treatment, we propose a model-based approach because it will increase the precision of those estimates.[39] The empirical model will be estimated with an OLS regression (using Stata).[40] We present the empirical model here:

$$Y_{Post} = \beta_0 + \beta_1 T + \beta_2 Y_{Pre} + \sum(\beta_P X_P) + \varepsilon$$

Where:
Y_{Post} – The outcome variable of interest, either: 1) times having sex without condoms in the past 3 months (continuous; values range 0 to *k*, where 0= has had sex without condoms 0 times in past 3 months, and *k*= number of times having sex without condoms in past 3 months); or 2) number of sexual partners in the past 3 months (continuous; values range 0 to *k*, where 0= has 0 sexual partners in past 3 months, and *k*= number of sexual partners in past 3 months) reported by participant *i* at the 9-month follow-up. (see Table 1 for full details on the variable construction).

Y_{Pre} – The baseline measure of the outcome variable of interest reported by participant *i* at baseline (see Table 2 for full details on the variable construction); variable will be re-centered at the grand mean for analysis.

[39] With the assumption that we maintain low attrition and differential attrition and that the study otherwise executes the RCT with integrity, we should be able to estimate an un-biased estimate of the average treatment effect of the intent to treat participants with *PS-R* by comparing differences in the means of the outcome variable reported by the treatment group with those reported by the control group. We could then provide a compelling response to our research question by testing the hypothesis that there is no difference between the two groups using straight-forward hypothesis testing statistics (t-test). With that said, we propose a regression-based model that includes covariates, because randomization should ensure covariates are uncorrelated with the treatment variable (i.e., they should not affect the estimate of the treatment effect), and in the instance they are significant predictors of the outcome, their inclusion in a regression model will decrease the standard error of the estimates, making them more precise. See: Angrist, J. D., & Pischke, J. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton: Princeton University Press; Rosenblum, M. and van der Laan, M. J. (2009), Using Regression Models to Analyze Randomized Trials: Asymptotically Valid Hypothesis Tests Despite Incorrectly Specified Models. Biometrics, 65: 937-945. doi:10.1111/j.1541-0420.2008.01177.x.

[40] As part of our sensitivity analyses, we will construct a logistic regression model to explore any potential differences in our effect estimates given the model utilized.

T –A dummy treatment indicator variable whose value equals 1 if the participant is randomized into the treatment group and zero otherwise.

X – A p vector of baseline (i.e., measured prior to receiving intervention or exogenous to treatment) participant-level covariates as well as blocking variables to account for the variation in outcomes associated with these groups. These covariates, listed in detail in Table 2 in the appendix, will include:

a) Age – age (based on date of birth) reported by agency staff or therapist at screening (continuous; range 14-19); variable will be re-centered at the grand mean for analysis.
b) Race – race of participant as self-reported at baseline. Race will be coded as a set of 4-1 = 3 dummy variables (each coded as 1 if they are of the specified race and coded as 0 otherwise); each of the variables will be re-centered at the grand mean for analysis.
c) Hispanic/Latino – self-reported as Hispanic, Latino, or of Spanish origin at baseline (0=do not identify as Hispanic/Latino/Spanish origin; 1=identify as Hispanic, Latino, of Spanish origin); variable will be re-centered at the grand mean for analysis.
d) High school education – self-reported education at baseline. A dummy variable (0= has not completed high school; 1= has completed high school); variable will be re-centered at the grand mean for analysis.
e) State – An n-1 vector of state indicator dummy variables that are coded one if the intervention was delivered at state n and coded zero otherwise. State 1 is the reference category and is excluded from analysis. The dummy variables will be mean-centered for analysis to facilitate interpretation.
f) Female – A dummy variable that is coded 1 if the participant self-reported being female at screening and coded 0 otherwise; variable will be re-centered at the grand mean for analysis.

β_0 – The intercept term, which represents, depending on the outcome measure of interest in the analysis, the outcome for the average control participant with all other variables in the model held constant at their mean.
β_1 – This is the parameter estimate of substantive interest. β_1 represents, depending on the outcome measure of interest in the analysis, either: 1) the adjusted mean difference in treatment and control participants' self-reported times having sex without condoms in the past three months at the nine-month follow-up; or 2) the adjusted mean difference in treatment and control participants' self-reported number of sexual partners in the past three months at the nine-month follow-up.
ε – The error term or unexplained individual-level variance that remains for each observation after the structural components of the model are estimated. It is the difference between the observed and the predicted values at the individual level.

We will report model-estimated effects and the results of significance tests in the findings section of the final impact report. Statistical significance will be based on test statistics produced by Stata for the coefficient β_1 using a two-tailed test, with $p < .05$.

ii. **Sample attrition**: Overall and differential attrition will be calculated using the full sample of participants enrolled in the study. This will be determined using data within the *Participant Database*. We will calculate overall attrition using the following formula:

$$Attrition_{participant} = 1 - \left(\frac{Assessed_{participant}}{Base_{participant}}\right)$$

where:

$Attrition_{participant}$ - the proportion of participants enrolled in the study who did not complete a 9-month follow-up questionnaire

$Base_{participant}$ - the number of participants enrolled into the study

$Assessed_{participant}$ - the number of participants who completed a 9-month follow-up questionnaire

Differential attrition will be calculated using the following formulas:

$$Attrition_{participant}\, C = 1 - \left(\frac{Assessed_{participant}C}{Base_{participant}C}\right)$$

$$Attrition_{participant}\, T = 1 - \left(\frac{Assessed_{participant}T}{Base_{participant}T}\right)$$

$$Differential\ Attrition_{t-c} = abs(Attrition_{participant}\, T - \ Attrition_{participant}\, C)$$

Where:

$Differential\ Attrition_{t-c}$ - the absolute difference between the proportion of treatment group participants who did not complete a 9-month follow-up questionnaire and the proportion of control group participants who did not complete a 9-month follow-up questionnaire

$Attrition_{participant}C$ - the proportion of participants enrolled in the study and randomly assigned to the control group who did not complete a 9-month follow-up questionnaire

$Base_{participant}C$ - the number of participants enrolled into the study and randomly assigned to the control group

$Assessed_{participant}C$ - the number of participants randomly assigned to the control group who completed a 9-month follow-up questionnaire

$Attrition_{participant}T$ - the proportion of participants enrolled in the study and randomly assigned to the treatment group who did not complete a 9-month follow-up questionnaire

$Base_{participant}T$ - the number of participants enrolled into the study and randomly assigned to the treatment group

$Assessed_{participant}T$ - the number of participants randomly assigned to the treatment group who completed a 9-month follow-up questionnaire

Overall and differential attrition will be reported in the findings section of our final impact report.

iii. **Adjustments for multiple comparisons:** Following guidance provided under the grant for our impact analysis plan[41], we will adjust for multiple comparisons in all of our primary

[41] During the January 8, 2019 OAH TPP Tier 2b Group Call on Impact Analysis Plans, the presenters noted that multiple comparison adjustment is required for all model-generated effect estimates of primary outcome measures.

outcome s analyses, regardless of outcome domains. We propose to use the Benjamini-Hochberg method.[42] This method controls for the false discovery rate (FDR), which is the expected value of the number of false positive tests divided by the total number of significant tests within a family of tests. The following procedures will be used to implement this adjustment:

1. The p-values generated by our models of the effect of the intervention on our three primary outcome measures will be ranked from smallest to largest, indexed by i (where $i = 1$ for the smallest p-value and $i = k$ for the largest p-value).
2. Beginning with the largest p-value (p_{k1}), we will assess if $p_{k1} < ((i/m)a^*)$, where m = the total number of tests conducted, and a^* = the initial significance value at which we would reject the null hypothesis and the level of false discovery we are willing to accept (in this case, 0.05). The null hypothesis will be rejected and the test will be considered statistically significant if $p_{k1} < ((i/m)a^*)$. If $p_{k1} < ((i/m)a^*)$, all smaller p-values in the list will also be considered statistically significant and the null hypothesis will be rejected for each test. If $p_{k1} \geq ((i/m)a^*)$, the null hypothesis will hold, the test will not be considered statistically significant, and the next largest p-value in the ranked list will be assessed.
3. If the 1st p-value is not statistically significant, the 2nd largest p-value in the list (p_{k2}) will be compared against $(i/m)a^*$. The null hypothesis will be rejected and the test will be considered statistically significant if $p_{k2} < ((i/m)a^*)$. If $p_{k2} < ((i/m)a^*)$, all smaller p-values in the list will also be considered statistically significant and the null hypothesis will be rejected for each test. If $p_{k2} \geq ((i/m)a^*)$, the null hypothesis will hold, the test will not be considered statistically significant, and the next largest p-value in the ranked list will be assessed.
4. If the 2nd p-value is not statistically significant, the 3rd largest p-value in the list (p_{k3}) will be compared against $(i/m)a^*$. The null hypothesis will be rejected and the test will be considered statistically significant if $p_{k3} < ((i/m)a^*)$. If $p_{k3} \geq ((i/m)a^*)$, the null hypothesis will hold and the test will not be considered statistically significant.

iii. **Sensitivity analyses:** We will conduct sensitivity analyses to test the robustness and validity of our benchmark approaches outlined above. These include: (1) excluding covariates; (2) not adjusting for missing baseline data; (3) excluding unreliable data; (4) excluding outliers; (5) condensing data collection windows to exclude late responders; and (6) estimating an empirical model that uses appropriate alternative approach if analytic sample size is not sufficient for benchmark approach.

1. **Without baseline covariates.** Our benchmark approach is to include baseline covariates in our model to improve the precision of our estimates. To test this, we will conduct sensitivity analyses that involve running identical empirical models without the covariates included. Analytic findings for both approaches will be presented alongside each other in an appendix of the impact report.

2. **Without adjusted baseline data.** As outlined in the *Missing data approach section*, our benchmark approach is to adjust baseline data as published guidance suggests that this may produce unbiased impact estimates and maximize the use of available data. We will test this by way of sensitivity analyses that involve running identical empirical models without the adjusted data. Analytic findings for both approaches will be presented alongside each other in an appendix of the impact report.

[42] This method has been selected because it helps to control the Type 1 error rate without also increasing the Type 2 error rate, which in our view is a serious consideration in preliminary efforts to identify evidence of effectiveness of new approaches. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological), 289-300.

As outlined in the *Baseline equivalency* section, we will also produce diagnostic estimates of baseline equivalency on the pre-intervention outcome variables according to our benchmark approach and the sensitivity study alongside each other in an appendix of the report.

3. **With unreliable data.** As discussed in the *Data cleaning* section, data for cases that are deemed unreliable are treated as *unit missing* and excluded from benchmark analyses. To test this, we will conduct sensitivity analyses that involve running identical empirical models with the unreliable data included. Analytic findings for both approaches will be presented alongside each other in an appendix of the impact report.

4. **Without outliers**. As discussed in the *Data cleaning* section, extreme data values are investigated and flagged as outliers. Our benchmark analytic approach is to include data flagged as outliers (i.e., extreme values that are not considered invalid) in analysis. We will also conduct sensitivity analyses that exclude these data and report substantive differences in the results section of the report.

5. **Condensed data collection windows**. Our benchmark approach is to include follow-up data from all participants who completed a questionnaire during their open data collection window, regardless of the time point in that window when it was completed. Data collection windows are broad to minimize attrition from the analytic sample. To examine whether or not this influences our results – and, in particular, whether or not study participants who respond later report different outcomes from those who respond earlier – we will conduct an analysis that examines the difference, if any, in response time between treatment and control participants and compares impact estimates for analytic samples without late responders. Late responders will be defined as those participants who complete their nine-month questionnaire more than one month after the initiation of the nine-month data collection window.

6. **Using alternate model specification**. Our benchmark approach is to use OLS regression to estimate the effects of the intervention on each of our primary outcomes of interest. This approach is deemed appropriate even in the instance the dependent variable is not normally distributed – as may be the case with our outcomes – so long as the sample size is sufficiently large.[43] In the instance our analytic sample size does not reach 500 participants (the threshold at which we consider a sample large), we propose to conduct diagnostic tests to identify an appropriate alternative estimation approach (e.g., transformation of the dependent variable, Bayesian approach, non-parametric model) that will be used to conduct sensitivity analyses to test the robustness of our results.

5. **Additional planned analyses:** We intend to investigate the following exploratory research questions in addition to the primary research questions described above.

<u>Antecedents of Sexual Behavior</u>

a. What are the short-term (post-program) and long-term (nine-month follow-up) impacts of the offer to participate in *PS-R* (treatment) relative to the offer to participate in *TPG (*comparison) on the following antecedents of participants' sexual behavior:
    i. Perception of risk and severity for pregnancy and HIV/STIs

43 Lumley, T., Dieher, P., Emerson, S., Chen, L. (2002). The Importance of the Normality Assumption in Large Public Health Datasets. *Annual Review of Public Health*. 23:151–69
DOI: 10.1146/annurev.publheath.23.100901.140546

ii. Intention to have sex
iii. Importance of having sex
iv. Intention to use effective contraceptive methods (including condoms)
v. Importance of using effective contraceptive methods
vi. Intention to limit number of sexual partners
vii. Importance to limiting the number of sexual partners
viii. Intention to practice affect regulation
ix. Importance of using affect regulation
x. Intention to practice sexual self-regulation
xi. Importance of practicing sexual self-regulation
xii. Contraceptive use and negotiation (including condoms) self-efficacy
xiii. Sexual activity decision making and planning self-efficacy
xiv. Beliefs regarding malleability of emotions
xv. Beliefs regarding value of affect regulation
xvi. Affect regulation self-efficacy
xvii. Use of affect regulation
xviii. Self-esteem
xix. Distress

Primary Outcomes Measured at Short-term Follow-up

a. What are the short-term (post-program) impacts of the offer to participate in *PS-R* (treatment) relative to the offer to participate in *TPG (*comparison) on the primary outcomes of interest:
    i. Number of times having sex without a condom
    ii. Number of sexual partners

Other Sexual Behaviors and Sexual Health Outcomes

a. What are the short-term (post-program) and long-term (nine-month follow-up) impacts of the offer to participate in *PS-R* (treatment) relative to the offer to participate in *TPG (*comparison) on the following sexual behaviors:
    i. Number of times having sex without any protection (prescription birth control or condoms)
    ii. Number of times having sex using dual methods of protection (prescription birth control and condoms)
b. What are the long-term (nine-month follow-up) impacts of the offer to participate in *PS-R* (treatment) relative to the offer to participate in *TPG (*comparison) on the following pregnancy outcomes
    i. Ever being pregnant?
    ii. Times being pregnant?

Effects of Mediators on Primary Outcomes of Interest

a. What are the short-term (post-program) and long-term (nine-month follow-up) impacts of the offer to participate in *PS-R* (treatment) relative to the offer to participate in *TPG (*comparison) on participants' reported times having sex without condoms considering the following potential mediators:
    i. Perception of risk and severity for pregnancy and HIV/STIs
    ii. Intention to have sex
    iii. Importance of having sex
    iv. Intention to use effective contraceptive methods (including condoms)
    v. Importance of using effective contraceptive methods
    vi. Intention to limit number of sexual partners
    vii. Importance to limiting the number of sexual partners
    viii. Intention to practice affect regulation

ix. Importance of using affect regulation
x. Intention to practice sexual self-regulation
xi. Importance of practicing sexual self-regulation
xii. Contraceptive use and negotiation (including condoms) self-efficacy
xiii. Sexual activity decision making and planning self-efficacy
xiv. Beliefs regarding malleability of emotions
xv. Beliefs regarding value of affect regulation
xvi. Affect regulation self-efficacy
xvii. Use of affect regulation
xviii. Self-esteem
xix. Distress

b. What are the short-term (post-program) and long-term (nine-month follow-up) impacts of the offer to participate in *PS-R* (treatment) relative to the offer to participate in *TPG (*comparison) on participants' reported number of sexual partners considering the following potential mediators:
   i. Perception of risk and severity for pregnancy and HIV/STIs
   ii. Intention to have sex
   iii. Importance of having sex
   iv. Intention to use effective contraceptive methods (including condoms)
   v. Importance of using effective contraceptive methods
   vi. Intention to limit number of sexual partners
   vii. Importance to limiting the number of sexual partners
   viii. Intention to practice affect regulation
   ix. Importance of using affect regulation
   x. Intention to practice sexual self-regulation
   xi. Importance of practicing sexual self-regulation
   xii. Contraceptive use and negotiation (including condoms) self-efficacy
   xiii. Sexual activity decision making and planning self-efficacy
   xiv. Beliefs regarding malleability of emotions
   xv. Beliefs regarding value of affect regulation
   xvi. Affect regulation self-efficacy
   xvii. Use of affect regulation
   xviii. Self-esteem
   xix. Distress

Appendix A: Tables

Table 1. Behavioral outcomes used for primary impact analyses research questions

Outcome name	Description of the outcome, including how it is operationalized	Source of the measure	Timing of measure
Times having sex without condom	The risk outcome is operationalized as the number of times in the past three months a person reports having any type of sex without using a condom. The measure is calculated from the following items: • In the past 3 months, how many times have you had vaginal sex without using a condom? • In the past 3 months, how many times have you had oral sex without using a condom? • In the past 3 months, how many times have you had anal sex without using a condom? The measure is calculated by summing the total number of times a person reported not using a condom during vaginal, oral and anal sex. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has not engaged in sex without a condom in the past three months, and *k* indicates the number of times the person has engaged in sex without a condom (risk behavior) in the past three months. Note: the analytic sample will include all respondents who have nine-month follow-up data. Persons who indicate they have not had sex will be considered to have participated in the risk behavior 0 times (i.e., they did not engage in sex without a condom).	*Participant Questionnaire*	Nine-month follow-up (nine months after the end of the 18-week intervention period)
Number of sexual partners	The risk outcome is operationalized as the number of sexual partners in the past three months. The measure is calculated from the following item: • How many sexual partners have you had in the past 3 months? The measure is calculated by summing the total number of sexual partners reported by the participant. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has had no sexual partners in the past three months, and *k* indicates the number of sexual partners in the past three months.	*Participant Questionnaire*	Nine-month follow-up (nine months after the end of the 18-week intervention period)

	Note: the analytic sample will include all respondents who have nine-month follow-up data. Persons who indicate they have had no sexual contact will be considered to have 0 sexual partners.	

Table 2. Covariates included in primary impact analyses

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion
Behavioral outcomes at baseline		
Times having sex without condoms	The risk outcome is operationalized as the number of times in the past three months a person reports having any type of sex without using a condom. The measure is calculated from the following items: • In the past 3 months, how many times have you had vaginal sex without using a condom? • In the past 3 months, how many times have you had oral sex without using a condom? • In the past 3 months, how many times have you had anal sex without using a condom? The measure is calculated by summing the total number of times a person reported not using a condom during vaginal, oral and anal sex. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has not engaged in sex without a condom in the past three months, and *k* indicates the number of times the person has engaged in sex without a condom (risk behavior) in the past three months.	Times having sex without condoms is included in the primary impact analysis as the pre-intervention or baseline measure of the behavioral outcome; it is included in the models so that individual-level change or difference can be assessed at the nine-month follow-up.
Number of sexual partners	The risk outcome is operationalized as the number of sexual partners in the past three months. The measure is calculated from the following item: • How many sexual partners have you had in the past 3 months? The measure is calculated by summing the total number of sexual partners reported by the participant. The resulting variable is continuous with values that range from 0 to *k*, where 0 indicates that a person has had no sexual partners in the past three months, and *k* indicates the number of sexual partners in the past three months.	Number of sexual partners is included in the primary impact analysis as the pre-intervention or baseline measure of the behavioral outcome; it is included in the models so that individual-level change or difference can be assessed at the nine-month follow-up.
Individual level covariates		
Age	The variable is measured as the respondent's age in years at screening. The measure is constructed from the following item on the *recruitment and eligibility form*:	Research has shown that likelihood of engaging in sex increases with age, while use of condoms declines and number of sexual partners increases

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion
	• Date of birth The variable is calculated by subtracting the reported date of birth given from the date when the screening was completed. The resulting variable is continuous with values ranging from 14 to 19.	(brewster 1999; kirby 2007; miller et al 1998; scott-jones and white 1990)
Race	The measure is operationalized a set of n-1 dummy variables, where *n* refers to the categorized race. The measure is taken from the following item on the baseline *Participant Questionnaire:* • What is your race? (Participants can select more than one response) o White o Black or African American o Hispanic, Latino, or of Spanish origin o American Indian or Alaska Native o Asian o Native Hawaiian or Pacific Islander o Some other race (specify) The category *Hispanic, Latino, or of Spanish Origin* is not considered in the operationalization of the race variable (it is used to create a separate Hispanic/Latino indicator variable). Remaining responses are recoded into the following mutually exclusive categories and dummy variables are created for each: *White only, Black only, Other race only* (which includes *American Indian/Alaskan Native, Asian, Native Hawaiian/Other Pacific Islander,* and *Other*), *Race not selected* (individual did not select any racial category), and *Multiracial* (which includes individuals who selected more than one racial category). Each dummy will be coded as 1 if the individual is coded as that particular race and 0 otherwise.	Research has shown that Black/African American and Hispanic adolescents are more likely to engage in sex during adolescence and initiate sexual activity at a younger age (Blum 2000; Brewster 1999; Hogan et al 2000; Kirby 2007; Scott Jones and White 1990)

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion

Hispanic, Latino or of Spanish origin	The measure is operationalized as a dummy variable, where 0 = identify as another ethnicity/do not identify ethnicity; 1 = identify as Hispanic, Latino or of Spanish origin.
	
	The measure is taken from the following item on the baseline *Participant Questionnaire:*
	
	• What is your race? (Participants can select more than one response)
	  ○ White
	  ○ Black or African American
	  ○ Hispanic, Latino, or of Spanish origin
	  ○ American Indian or Alaska Native
	  ○ Asian
	  ○ Native Hawaiian or Pacific Islander
	  ○ Some other race (specify)
	
	Variable will be coded as 1 if participant self-identified as *Hispanic, Latino or of Spanish origin*, regardless as to whether other races/ethnicities are specified; Hispanic, Latino or of Spanish origin is not selected, the response will be coded as 0.
		Research has shown that Black/African American and Hispanic adolescents are more likely to engage in sex during adolescence and initiate sexual activity at a younger age (Blum 2000; Brewster 1999; Hogan et al 2000; Kirby 2007; Scott Jones and White 1990)

Covariate	Description of the covariate and how it will be used as a covariate in the analysis	Rationale for inclusion
High school education	This measure is operationalized as a dummy variable, where individuals with a high school education are coded as 1 and all others are coded as 0. The measure is taken directly from the following item on the baseline *participant questionnaire:* • What is the highest degree or level of school you have completed (if currently enrolled, select the previous grade completed or degree received)? ○ Grade 9-12 (please specify grade) ○ Regular high school diploma ○ Ged or alternative credential ○ Some college credit, but less than 1 year of credit ○ 1 or more years of college credit, no degree ○ Associate's degree ○ Bachelor's degree ○ Other (specify) Participants who select *grade 9-12* will be coded as 0=has not completed high school; all others will be coded as 1=has completed high school.	
Blocking covariates		
State	The measure is operationalized a set of n-1 dummy variables, where *n* refers to the number of the states over the course of the evaluation period. Data for the measure are obtained from the *enrollment log database.* Each dummy will be coded as 1 if the individual is enrolled in a particular state and 0 otherwise. State 1 is the reference variable. Dummy variables will be grand mean centered so that the intercept will then reflect the un-weighted mean site effect.	Randomization is blocked by state.
Gender	The measure is operationalized as a dichotomous variable, where 1 indicates that the participant enrolled is female and 0 if otherwise. Data for the measure are obtained from the *enrollment log database.*	Randomization is blocked by gender.

Table 3: Data Editing Rules

The following table provides PRG's general rules for editing data based upon responses given.

Category	**Data editing rule**
No response given to an item (coded as .f)	If data from a related variable can be used to infer a value, data will be logically edited. Otherwise, the value will be left as missing.
Invalid items (coded as .i)	Adjust missing baseline values
Outlying items (Outlier indicator variable coded as 1)	Keep in benchmark analysis; run sensitivity analyses excluding outliers
Inconsistent across-time items (coded as .k)	Adjust missing baseline values
Unreliable cases (Unreliable indicator variable codes as 1)	Exclude case from benchmark analysis; run sensitivity analyses including unreliable cases

Appendix B: Evaluation Abstract Submitted to OAH in September 2017

EVALUATION ABSTRACT: THE EVALUATION OF PRACTICE SELF-REGULATION IN CALIFORNIA, MAINE, MICHIGAN, NEW MEXICO, AND LOUISIANA

Grantee

Grantee Name: The Policy & Research Group
Project Lead: Lynne Woodward Jenner
Email address: ljenner@policyandresearch.com

Evaluator

Evaluator's Organization: The Policy & Research Group
Evaluator Lead: Eric Jenner
Email address: ejenner@policyandresearch.com

Intervention Name

Practice Self-Regulation (PS-R)

Intervention Description

PS-R is a manualized therapy intervention that is trauma-focused and performed one-on-one with youth ages 14 to 19 years old who are receiving individual outpatient counseling services. The intervention aims to decrease risky sex behaviors by increasing knowledge of sexual health and the impact of trauma on sexual decision making, readiness to change risky sexual behavior, ability to manage impulsive behavior, confidence in ability to negotiate safe sex, and intentions to practice safe sex. It also enables participants to explore the effect of previous trauma on their behavior more generally.

The intervention has three key components: (1) intensive four-day facilitator training provided by the developer, using the Trauma Outcome Process Workbook for Sexual Health Facilitator's Guide; (2) the Trauma Outcome Process Workbook for Sexual Health, which the youth works through on their own between sessions; and (3) ten one-on-one therapy-education sessions led by a trained facilitator, in which the facilitator covers sexual health education and facilitates a discussion of the workbook chapter the youth has completed. The content of the intervention reflects the seven essential topics outlined in the National Sexuality Education Standards (anatomy and physiology, puberty and adolescent development, identity, pregnancy and reproduction, sexually transmitted infections and human immunodeficiency virus (HIV), healthy relationships, and personal safety). The workbook contains self-directed activities that help simplify the complex concepts associated with trauma to help youth easily understand and apply effective coping strategies for self-regulation and optimal sexual decision making. In addition to providing these core components of sexual health education, this intervention provides a setting for discussion between the youth and facilitator about the effect of trauma, personal goals, beliefs, values, and choices that affect a person's sexual health and well-being. The intervention also uses components of expressive therapy, such as multisensory activities, to enhance the therapeutic experience for the youth and enable them to envision pathways toward reducing personal risk.

The intervention should be delivered in a private space (for example, a clinician's office) ideally over 10 consecutive weeks (one session each week), however, it might take longer depending on a youth's mental health needs and scheduling (for example, vacations or missed appointments). For the purposes of the study, the intervention window has been defined as an 18-week period. During the intervention window, intervention participants are expected to receive all 10 sessions of PS-R in place of their regularly scheduled counseling sessions with their therapist. However, at the therapists' discretion, the PS-R sessions can temporarily halt to address pressing or acute needs of the study participants. Although there are no explicit educational requirements for PS-R facilitators, to participate in the study, facilitators must be: (1) clinicians with advanced degrees (for example, a master of social work) and an active clinical license; (2) trained in study procedures by a PRG research analyst; and (3) participate in a four-day PS-R training, conducted by Joann Schladale.

Comparison Condition

Business as usual

Comparison Condition Description

The comparison condition will be business as usual; that is the typical therapy or counseling that participants would receive from facilitators. There will be no alternative program or additional activities offered to the participants assigned to the comparison group. The comparison condition is referred to as Therapy Practice Group for the purposes of the study.

Therapy models typically used by facilitators will vary, but it is likely that other trauma-informed interventions or therapies will be used with control participants (e.g., Trauma Focused Cognitive Behavioral Therapy). Facilitators have been instructed they can answer specific questions related to sexual health in therapy sessions, however, they are to provide information only and referrals for additional services when necessary as they would under typical circumstances. Partner agency representatives and participating private practitioners have confirmed

that no other curriculum-based teen pregnancy prevention programs or similar therapy-based sexual education programs will occur during the study period.

Behavioral Outcomes

Condom use and number of sexual partners

Non-behavioral Outcomes

Sexual health knowledge (related to pregnancy, HIV and sexually transmitted infection transmission, and methods of protection), intention to engage in sex and safe sex behaviors, sexual attitudes (related the value or importance of engaging in sex and safe sex behaviors, self-efficacy to practice safe sex (use condoms or contraceptives, negotiate condom use, make healthy and safe sexual decisions), affect regulation beliefs and attitudes (related to the malleability of emotions and value of affect regulation), self-efficacy to regulate affect, self-esteem, self-regulation tendencies (related to rational and impulsive decision making), practice of affect regulation (related to frequency and consistency with which youth regulate their emotions)

Sample and Setting

The study is being conducted in California, New Mexico, Michigan, Maine, and Louisiana. Staff will screen all youth receiving treatment or counseling services from study agencies or clinics and partnering private practitioners' offices for eligibility. The intent-to-treat sample will be comprised of eligible youth who are enrolled into the study during the two-year implementation period (July 1, 2016 to June 30, 2018). To be eligible, participants must (1) consent or assent to participate; (2) be at least 14 years old, but no older than 19 years old; (3) be appropriate for the study as deemed by agency staff; and (4) be receiving individual outpatient counseling services at one of the study's implementation sites. The exclusion criteria are: (1) previous participation in the Teen Health Study; (2) previous use of the Trauma Outcome Process workbook; (3) self-report of previous participation in other teen pregnancy prevention-funded programs; or (4) self-report of roommate who participated in the Teen Health Study. The evaluation plans to enroll 600 eligible youth over the course of the study period.

Research Design and Data Collection

The study is an individual randomized controlled trial in which evaluators randomly assign eligible, consenting participants to intervention or comparison conditions at a one-to-one ratio. Random assignment occurs after participants' evaluation consent or assent and before the provision of any programming or collection of baseline data.

Randomization is blocked at the regional (state) level, which does not guarantee in each site that facilitators will deliver both the PS-R and comparison conditions; however, the expectation is that over the course of the implementation period facilitators will likely deliver both. Although this minimizes potential therapist-related confounds, it increases the risk of comparison group contamination. To reduce contamination risk, facilitators will train on the expectations of the study; following each session with comparison group participants, they will be asked to self-report the type of therapy provided and whether core components of PS-R were covered in the session. In addition, every therapist will be instructed to video record two randomly selected comparison condition sessions for each participant who gives permission to video record; a team of observers picked by the intervention developer will review these sessions to assess whether comparison youth are exposed to the intervention.

Participants in both the intervention and comparison groups will be surveyed at baseline (before the first intervention session), 18-week follow-up (post-baseline) survey, and 9-month follow-up (post-18-week follow-up). Participants will complete questionnaires (available in English and Spanish) electronically on a computer in a private room; however, paper questionnaires will be available if participants require them. The preferred mode for all data points is in-person data collection, but evaluators will collect data online, by mail, or over the phone in an interview format with difficult-to-reach participants at 18-week and 9-month follow-ups. Data collection windows for both follow-up questionnaires will be four months. The evaluators will track when participants respond and sensitivity analyses will examine whether outcomes differ for participants who responded on time (within the first month of the data collection window) and late (in the second, third, or fourth months of the window).

For the implementation evaluation, the evaluators will collect data on fidelity, attendance, and quality. Evaluators will collect these data through attendance logs and therapist reports (updated and completed after every session), as well as video observation reports (completed for 10 percent of all intervention and comparison participant sessions).

Schedule/Timeline

Sample enrollment and baseline data collection began July 2016 and will end December 2018. The 18-week post-baseline data collection began November 2016 and will end September 2019, and the 9-month follow-up data collection began August 2017 and will end June 2020.

Prepared for the Office of Adolescent Health by Mathematica Policy Research under contract #GS-10F-0050L, Task Order No. HHSP233201300416G with Policy and Research, LLC with grant #6 TP2AH000043.