

**Findings from the
Replication of an
Evidence-Based
Teen Pregnancy
Prevention
Program**

Evaluation of Becoming a Responsible Teen in New Orleans, LA

Final Impact Report for

Louisiana Public Health Institute

September 2015

Prepared by

The Policy & Research Group

Eric Jenner, PhD

Sarah Walsh, PhD

Lynne W. Jenner, MA

Hilary Demby, MPH

Alethia Gregory, LMSW

Erin Davis, BS

Recommended Citation: The Policy & Research Group (2015). *Evaluation of Becoming a Responsible Teen: Findings from the Replication of an Evidence-based Teen Pregnancy Prevention Program*. New Orleans, LA: Jenner, E., Walsh, S., Jenner, L. W., Demby, H., Gregory, A., & Davis, E

Acknowledgements:

The Policy & Research Group (PRG) would like to acknowledge the work of the Louisiana Public Health Institute (LPHI) Orleans Teen Pregnancy Prevention Project staff in collaborating with PRG to meet the stringent study requirements and overseeing the implementation of the *4 Real Health – Health Education Program* (HEP) during three summers. This effort was led by the LPHI Project Director, Dr. Marsha Broussard, along with Jenny Dickherber and Shantice Atkins. The City of New Orleans Office of Workforce Development was an essential partner, working with LPHI and PRG each summer to integrate HEP into the NOLA Youth Works youth job training program; Nadiyah Coleman and Tiffany Henderson were particularly instrumental in this task as were the community-based organization partners that hosted the program each summer. The Institute of Women and Ethnic Studies (IWES) conducted fidelity monitoring. And, finally, a cadre of Health Educators, trained by IWES and LPHI, who implemented the two HEP interventions (Becoming a Responsible Teen and Healthy Living).

This publication was prepared under Grant Number 5 TP1AH000003-02-00 from the Office of Adolescent Health, U. S. Department of Health & Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the policies of HHS or the Office of Adolescent Health.

EVALUATION OF BECOMING A RESPONSIBLE TEEN IN NEW ORLEANS, LOUISIANA: FINDINGS FROM THE REPLICATION OF AN EVIDENCE-BASED TEEN PREGNANCY PREVENTION PROGRAM

I. Introduction

A. Introduction and study overview

In 2009, the U.S. Department of Health and Human Services (HHS) began a systematic review of the evidence on programs designed to reduce and prevent teen pregnancy and sexually transmitted infections (STIs).ⁱ Although over 35 programs have been identified as having a statistically significant positive effect on sexual behaviors (e.g., frequency of sexual activity, number of sexual partners, use of contraception) or related outcomes (e.g., pregnancy, STIs), the review has also identified the absence of replication studies as a primary weakness in the research.ⁱⁱ In light of this, HHS incorporated replication studies into the Office of Adolescent Health (OAH) Teen Pregnancy Prevention Program (TPP) funding as a means to bolster the current evidence base.ⁱ

This report describes an evaluation of Becoming a Responsible Teen (BART), one of the programs identified by the HHS review as having a positive effect on sexual behaviors. BART is an out-of-school, group-level, cognitive behavioral and skills training sexual education course designed to reduce African American adolescents' risk for contracting HIV. The authors of an individual-level randomized controlled study of BART report that it evidenced a positive statistically significant impact on social-cognitive behavioral antecedents to safer sex behaviors (e.g., knowledge, attitudes, and self-efficacy), increasing behavioral and social skills related to the practice of safe sex (e.g., ability to handle coercive situations), and reducing or delaying some sexual risk behaviors (e.g., engagement in sex, unprotected sex).ⁱⁱⁱ The HHS evidence

review reports that the study received a “high” rating but found that only some of these outcomes met review standards or fit in the scope of the review.

In 2010, the Louisiana Public Health Institute (LPHI) received a five-year TPP Replication of Evidence-Based Programs (Tier 1) grant to replicate and rigorously evaluate the effectiveness of an evidence-based TPP program in New Orleans, Louisiana. Louisiana teen birth rates and rates of STIs, especially among African Americans, are among the highest in the country; in addition, there is a high prevalence of HIV in the state.^{iv v} LPHI selected BART because it was designed for use with African American adolescents and because it is appropriate for use with teens who are sexually experienced, as well as those who are not, and the program can be delivered to both males and females.

LPHI contracted with The Policy & Research Group (PRG), an independent research firm, to conduct the evaluation. This report presents findings from an implementation evaluation and a rigorous, high-quality randomized controlled trial (RCT) of the impacts of BART on two self-reported sexual behaviors (condom use and frequency of sex) at six-month follow-up.

B. Primary research question

Our primary research question is: What is the impact of the offer to participate in BART (treatment) relative to the offer to participate in Healthy Living (control) on participants’ reported inconsistent use of condoms six months after the end of treatment?

C. Secondary research question

Our secondary research question is: What is the impact of the offer to participate in BART (treatment) relative to the offer to participate in Healthy Living (control) on participants’ reported frequency of sex six months after the end of treatment?

II. Program and comparison programming

This report presents the results of an RCT in which eligible and consenting individuals were randomly assigned to one of two study conditions. The treatment condition is BART, an out-of-school, social-cognitive behavioral and skills training sexual education course designed to reduce African American adolescents' (ages 14 to 18) risk for contracting HIV. The control (counterfactual) condition is Healthy Living, a general health education program that aims to improve adolescents' nutrition, healthy eating habits, body image, and increase exercise. Like other programs that have been developed to reduce sexual risk behaviors in the past 20 years, BART is based in theory that suggests that imparting knowledge alone is insufficient to reduce risk behaviors; instead behavior change must be motivated by socio-cognitive antecedents of behavior including skills, beliefs, attitudes, intentions, and self-efficacy to engage in protective behavior.^{vi vii viii ix x} Consequently, and consistent with the earlier study that is being replicated, the counterfactual experience includes the provision of factual information deemed necessary to prevent HIV and unintended pregnancy.

For the purposes of study, BART and Health Living were referred to jointly as *4 Real Health – Health Education Program* (HEP). In collaboration with its partners, the City of New Orleans and the Institute for Women and Ethnic Studies (IWES), LPHI implemented HEP over three consecutive summers (2012-2014) in New Orleans, LA as an educational component of a youth summer employment program, NOLA Youth Works (NYW), funded by the city government. NYW contracts with multiple community-based organizations (CBOs) to offer summer camps, internships, job training, and employment opportunities for youth ages 13 to 21 who reside in Orleans Parish. To be eligible to participate in the study, NYW sites were expected to meet pre-established criteria including having the capacity to host at least 40 youth for at least six consecutive weeks and having a minimum of two separate classrooms available for

programming. LPHI project staff coordinated with NYW and their partners to secure programming space each summer and to manage day-to-day operations of BART and Healthy Living at the NYW sites. In all, 12 CBOs participated as partners, and implementation was carried out at 18 work sites (several CBOs implemented the program at multiple sites).

LPHI hired health educators to administer HEP and partnered with IWES to conduct curriculum training and fidelity monitoring for the program. IWES and LPHI trained health educators in both BART and Healthy Living interventions and provided the following supplemental trainings: HIV 101, Cultural Competency, Nutrition Basics, Mandatory Reporting, Classroom Management, Fidelity Monitoring, and Evaluation Research Basics. Health educators worked in teams (one male and one female) to deliver HEP. To minimize instructor effects, each team was expected to teach both interventions to classes of the same gender. For instance, if a team taught Healthy Living to a female class, they would be expected to teach BART to a female class (or vice versa).

A. Description of program as intended

Developed within the context of social learning and self-efficacy theories, BART includes 4 core components—information, skills training, opportunities to practice skills, and social support—that are meant to increase participants’ knowledge and awareness of risk, clarify participants’ values related to sexual behaviors, develop and enhance participants’ risk reduction skills, build attitudes supportive of condom use, and foster intentions to reduce high-risk behaviors. By addressing these theoretically relevant motivational antecedents of sexual behavior, the program ultimately aims to increase safer sex behaviors of participants (e.g., increase consistency of condom use and reduce the frequency of sex) thereby preventing the transmission of HIV among African-American youth and reducing teen pregnancy.

Program content is designed to be delivered in eight sessions over the course of eight weeks. Each session is expected to take 90 to 120 minutes. BART fidelity requirements mandate that the intervention be delivered in small gender-specific groups of between 5-15 persons by two co-leaders, one male and one female. The course curriculum begins with an introductory HIV/AIDS informational session intended to increase awareness of risk and dispel common HIV myths (session 1); subsequent sessions are directed at making decisions and clarifying values (session 2), building condom use and assertive communication skills (sessions 3-5), personalizing risk (session 6), and understanding the importance of sharing course content with others (sessions 7 and 8). Throughout the course, participants are presented with facts about HIV transmission, risks, and prevalence, as well as facts about how to protect themselves from risk. Although abstinence is presented as the best protection against sexual risks, the curriculum primarily focuses on the development of condom use and negotiation skills. Table II.1 below presents an overview of the curriculum, along with the number of activities involved in each session.

For this study, no adaptations to program components or content were planned prior to the start of implementation. However, following the first summer of implementation (September 2012), LPHI received approval to reduce program duration from eight weeks to six weeks. This adaptation did not alter the BART curriculum content or reduce the total number of sessions offered (during two program weeks, two sessions were offered instead of one session).

Table II.1. Intended Program Content for BART, by Session

Session	Overview	Number of Activities
1. Understanding HIV & AIDS	Provides information on what HIV is, how it is transmitted, risk and protective behaviors, and HIV prevalence among the target population; it also dispels common HIV myths	6
2. Making sexual decisions and understanding your values	Reviews information on HIV transmission, risks, stereotypes, and prevalence; it also includes activities intended to personalize risk and to help participants identify support systems	7
3. Developing & Using Condom Skills	Presents facts about condoms, examines attitudes toward condoms and common barriers to their use, and provides demonstration of how to use condoms	5
4. Learning Assertive Communication Skills	Presents ways to negotiate safer sex, identifies common communication problems and possible solutions, and demonstrates different communication styles	4
5. Practicing Assertive Communication Skills	Presents tips for assertive communication, explores ways to say no; demonstrates and allows participants to practice assertive communication through role-play	5
6. Personalizing the Risks	Presents personal accounts of HIV through in person presentations or videos	2
7. Spreading the Word	Participants link assertive communication skills to their lives and identify ways to get out of risky situations; demonstrates and allows participants to practice sharing what they have learned	4
8. Taking BART with You	Reviews HIV facts; participants discuss how their behaviors or attitudes have changed; experiences sharing what they learned	4

B. Description of counterfactual condition

The counterfactual condition (Healthy Living) was designed as a general health and nutrition course, with dosage and implementation requirements identical to BART. Specifically, Healthy Living was intended to be delivered in eight 90-120 minute sessions over the course of eight weeks; it was to be delivered in small, gender-specific groups of 5 to 15 youth, and it was to be facilitated by teams of two health educators (one male and one female).

The first session of Healthy Living was intended to be identical to the first session of BART – that is, it offered the same initial HIV information-only session as BART. This session provides participants with information on who is at risk for HIV and why, HIV terms, common HIV facts and myths, HIV risk behaviors, and how to use knowledge about HIV to positively

influence others. Sessions 2 through 8 were adapted from the Oregon Dairy Council's "Live It! Real-Life Nutrition for Teens" curriculum and included information and activities related to basic nutrition and dietary guidelines, healthy food and healthy eating habits, body image, and physical activity and exercise. These sessions were to be focused strictly on health and nutrition and were not to contain any sexuality education components or to incorporate core elements of BART. In their HEP training for Healthy Living sessions 2-8, health educators were instructed not to discuss ways to handle social and sexual pressures, ways to communicate assertively about sex, refusal and negotiation skills related to sex, or condom use skills.

III. Study design

A. Sample recruitment

Participants for the study were recruited over three consecutive summers (2012-2014) from a youth summer employment program (NYW) in New Orleans, LA. To be eligible to participate, youth had to (1) be between the ages of 14 to 18, (2) be assigned to a NYW job site that offered HEP, (3) not have previously participated in a specified list of other OAH-funded TPP initiatives operating in Louisiana, and (4) provide parental consent (if under age 18) and participant assent to participate in the study.

There were no differences in the recruitment process for the treatment or control groups. Each year, evaluation staff recruited potential participants and screened for eligibility using consent packages that were provided to potential study participants and their parents (i.e., youth who had been accepted into the summer employment programs and could be placed at sites administering the interventions). Consent packages contained a cover letter explaining the study, parent program and evaluation consent forms (separate documents), and youth evaluation assent forms. Assent forms included eligibility questions related to the potential participants' gender, age, and any prior participation in other teen pregnancy prevention programs operating in the

city. If consent packages were not completed and returned to the evaluators, evaluation staff attempted to verbally assess eligibility and gain consent via phone or in person prior to the beginning of programming.

In all, 1,230 youth were assigned to work sites implementing HEP over the course of the evaluation. Of these, 959 provided consent/assent and were otherwise eligible for the study. The remaining 271 youth were not eligible to participate in the study because they did not provide consent, they had previously participated in a program funded by OAH TPP, or they were not of the correct age. Of the 959 youth who were eligible, 850 were enrolled and randomized into the study. The remaining 109 youth were not enrolled because they did not attend the work site during the randomization period.

B. Study design

The evaluation involves: 1) an individual-level randomized controlled trial (RCT) to assess the impact of BART on sexual behaviors; and 2) an implementation study to assess the fidelity and quality of programming that is intended to provide context for the impact findings. For the impact study, the evaluators randomly assigned individual participants prior to the collection of baseline data and the provision of programming. Various NYW program constraints (participants were assigned to different sites and shifts) and fidelity requirements for BART (which required gender-specific classes containing between 5-15 participants), necessitated a blocked design. Individuals were randomly assigned to treatment or control conditions within cohort and site, according to gender and work shift. That is, each summer, youth at each employment site were randomized into gender-specific treatment and comparison groups according to whether they worked in the morning or afternoon.

All individuals who were present during the first day of work at sites where HEP was implemented and who met all the eligibility criteria were individually randomly assigned into an

intervention or comparison group using the random allocation (ralloc) command in Stata. In addition, those youth who attended work at a HEP implementation site for the first time during the first or second week but had not been randomized due to their absence on the first day were randomized into the study on a rolling basis, provided they met the eligibility criteria and there was space in the class. The randomization procedures were slightly different for these individuals (the evaluators used a coin toss), but the probability of assignment to treatment group ($p = .50$) was equal in expectation to the ralloc procedure.

C. Data collection

1. Impact evaluation

To assess impacts on condom use and sexual activity for the impact study, baseline, outcome, and covariate data were collected via a self-administered questionnaire that was scheduled at baseline (before the first program session was attended) and six months postprogram. Extended data collection windows were used for each survey administration in order to maximize the amount of time individuals had to respond to the questionnaire. Baseline data collection windows were open for up to eight weeks (they closed just prior to the end of programming) and six-month follow-up data collection windows were open for up to six months. In Appendix E, we describe and present results from a sensitivity analysis conducted to test whether our results were affected by these methodological decisions.

Though participants were encouraged to complete the questionnaire in person, to accommodate busy schedules and capture initial non-responders, all participants (both treatment and control) were also offered the opportunity to complete the questionnaire via other modes (online, mail, phone interview) at specified time points within the data collection period. Incentives were provided to all study participants for completing each questionnaire, regardless as to mode of administration. In order to reduce attrition, evaluation staff developed extensive

follow-up protocols based on the Engagement, Verification, Maintenance, and Confirmation follow-up model. Protocol activities started during program implementation periods and extended throughout the six-month follow-up data collection periods; they were intended to engage participants, collect and verify contact information, maintain contact with participants post-program, keep contact information up-to-date, and schedule follow-up appointments.^{xi xii}

There were no differences in data collection efforts across the two assignment conditions. At each data collection point, the treatment and control groups were asked to complete the same questionnaire; the data collection schedule and variations in mode of administration were offered identically across groups, and both groups were offered the same incentives to participate in data collection (for a summary of data collection procedures and a detailed data collection timeline for each HEP cohort, see Tables A.1 and A.2, respectively, in Appendix A). Furthermore, inspection of administrative and participant data show that there were no substantive differences between treatment and comparison groups in terms of mode or timing of survey administration. Data presented in Table A.3 show that similar proportions of participants from both groups took the questionnaire in person, online, via mail, and by phone interview; in addition, as can be seen in Figures A.1 through A.3, the two groups completed baseline and six-month data collection in comparable lengths of time.

2. Implementation evaluation

The following data and data sources were used to assess the extent of programming offered and received as well as the fidelity and quality of program implementation: 1) sessions offered and received, content delivered to youth, and health educator background details (for both treatment and counterfactual) were collected using class attendance sheets, health educator and observer fidelity tools, and health educator administrative data, respectively; 2) overall quality of program sessions and delivery of information were collected by observer fidelity monitors using

the *Program Observation Form for TPP Grantees*; and 3) implementation context data were collected with two items on the participant questionnaire and administrative program documents. See Table B.1 in Appendix B for detailed implementation data sources, data collection frequency, and party responsible for data collection for each implementation element.

D. Outcomes for impact analyses

Our primary research question asks whether the offer to participate in BART relative to the offer to participate in Healthy Living impacts participants' inconsistency of condom use six months after the end of the intervention. We operationalize inconsistency of condom use as a risk outcome – the proportion of times in the past three months a participant *does not* use condoms while engaging in any type of sex. 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. Persons who indicate that they are not sexually active are considered to have engaged in the risk behavior 0% of the time.

Our secondary research question asks to what extent the offer to participate in BART relative to the offer to participate in Healthy Living impacts participants' reported frequency of sex (sexual activity) six months after the end of treatment. Our measure of frequency of sexual activity is a count variable – the self-reported number of times in the past three months a person engages in any type of sex. As with our primary impact analysis, in our assessment of this secondary outcome, we consider the self-reported sexual behaviors of all participants who report sufficient data, including individuals who indicate they are not sexually active. Persons who indicate that they are not sexually active are considered to have had sex zero times.

The evaluators infer that BART has the hypothesized impact on primary and secondary behavioral outcomes (inconsistency of condom use and frequency of sex) if, at six-month follow-up, participants assigned to BART report a (regression adjusted) mean outcome that is less than

that reported by participants assigned to Healthy Living – and the difference between the two means is significant. Statistical significance is determined at the $\alpha = .05$ level, using a two-tailed test. A detailed description the behavioral outcome measures used for primary and secondary impact analyses research questions are presented in Table III.1 (primary research question) and Table III.2 (secondary research question).

Table III.1. Behavioral outcomes used for primary impact analyses research question

Outcome name	Description of outcome	Timing of measure relative to program
Inconsistency of condom use	<p>The outcome is measured as the proportion of times in the past three months a person reports having any type of sex without using a condom.</p> <p>The outcome variable is calculated from the following items on the participant questionnaire administered 6 months after the program's end:</p> <ul style="list-style-type: none"> • In total, how many times have you had any type of sex in the past 3 months? • Now, think about the number of times that you had any type of sex in the past 3 months. How many of those times did you use condoms? <p>The number of times respondents did not use a condom is calculated by subtracting the number of times a person reports using a condom during sex from the total times s/he reports having sex.</p> <p>The outcome measure is calculated by dividing the total number of times a person reported not using a condom by the total number of times s/he reported having sex.</p> <p>The resulting variable is a continuous proportion with values that range from 0 to 1, where 0 indicates that a person has not engaged in sex without a condom in the past three months, and 1 indicates that the person has engaged in sex without a condom 100% of the times they had sex in the past three months.</p> <p>Note: Respondents are instructed that any type of sex refers to oral, anal, or vaginal sex, and not masturbation.</p>	6 months after program ends

Table III.2. Behavioral outcomes used for secondary impact analyses research question

Outcome name	Description of outcome	Timing of measure relative to program
Frequency of sexual activity	<p>The outcome is measured as the number of times in the past three months a person reports having any type of sex</p> <p>The measure is taken directly from the following item on the HEP Questionnaire</p> <ul style="list-style-type: none"> In total, how many times have you had any type of sex in the past 3 months? <p>The variable is continuous, with values ranging from 0 to k, where 0= no sexual activity reported in past 3 months and k = number of times sex reported.</p> <p>Note: Respondents are instructed that any type of sex refers to oral, anal, or vaginal sex, and not masturbation.</p>	6 months after program ends

E. Study sample

Table C.2 in Appendix C depicts the flow of sample members from the beginning of the study through the follow-up survey that was used to address the research questions. The full set of 850 participants who were offered the opportunity to participate in either BART or Healthy Living and who provided evaluation consent/assent constitutes the full intent-to-treat (ITT) sample; 427 youth were randomly assigned to receive the BART intervention and 423 were assigned to be control participants.

The analytic sample, which is the subset of the ITT sample for whom we have sufficient data, is 688 youth. Participants were considered to have sufficient data if they contributed reliable baseline and six-month follow-up questionnaires; if they contributed a questionnaire but did not provide a response to one or more questions used in the impact analyses their data were imputed (see the *Missing Data Approach* section of Appendix D). Of the 850 youth randomized, 111 (59 treatment and 52 control) were excluded from the study sample because they did not complete a baseline and/or six-month follow-up questionnaire, and 51 (26 treatment and 25 control) were excluded because at least one of their completed questionnaires was deemed

unreliable (see the *Data Cleaning Procedures* section in Appendix D for an explanation of what are considered unreliable data). Thus, 688 participants (342 treatment; 346 control) constitute the analytic sample for both the primary and secondary contrasts; this represents 80.9% of the full ITT sample. Baseline data collected on study participants indicate that just over half (53%) are female; nearly all identify as black (84%) or multiracial (15%), and a small percentage identify as Hispanic (3%). On average, participants were 15 years old at baseline and self-reported engaging in sex one time in the prior three months. Additionally, roughly three-quarters (72%) of the analytic sample report that they were not sexually active at the time of enrollment (i.e., they had not had sex in the three months prior) and just over half (55%) self-reported never having engaged in sex.

F. Baseline equivalence

We assess baseline equivalence of the treatment and control groups on pre-intervention measures of our primary and secondary outcomes (inconsistency of condom use and frequency of sex) and key covariates (age, gender, race, ethnicity, parental education, and family structure) for the analytic sample. We used a two-step procedure to establish balance. We first generated model-based estimates of the differences between groups and then examined the statistical significance of the differences. Separate models were run for each of the baseline variables. Ordinary least squares regression models were used to estimate differences in continuous baseline measures and linear probability models were constructed to estimate differences in dichotomous baseline measures. The first two columns in Table III.3 report descriptive statistics (regression adjusted means and unadjusted standard deviations) for the intervention and comparison groups separately. The final two columns report model coefficients (i.e., regression adjusted mean differences or predicted probability of group membership) and their associated p -

values. As can be seen in Table III.3, baseline equivalence is convincing. Differences between the treatment and control groups are small and statistically insignificant (i.e., $p > .05$ in all cases).

Table III.3. Summary statistics of key baseline measures for youth completing 6-month follow-up

Baseline measure	BART mean or proportion ^e (standard deviation)	Comparison mean or proportion (standard deviation)	Intervention versus comparison mean difference ^f	Intervention versus comparison <i>p</i> -value of difference
Age (years)	15.06(0.79)	15.03(0.85)	0.03	0.437
Gender (female)	0.25(0.5)	0.25(0.5)	0.00	0.938
Race: Black	0.78(0.36)	0.77(0.38)	0.01	0.628
Race: Multiracial ^a	0.22(0.35)	0.22(0.36)	0.00	0.912
Ethnicity: Hispanic	0.03(0.17)	0.04(0.19)	-0.01	0.656
Parental education level ^b	2.74(0.96)	2.74(0.9)	0.00	0.991
Family structure (lives with both parents)	0.20(0.44)	0.14(0.4)	0.05	0.101
Frequency of sexual activity ^c	1.13(2.77)	1.23(3.42)	-0.10	0.673
Inconsistency of condom use ^d	0.14(0.26)	0.15(0.27)	-0.02	0.343
Sample size	342	346		

(BART = Becoming a Responsible Teen)

Notes:^a Multiracial refers to individuals who selected more than one race category when asked “What is your race?”

^b Parental education level refers to the mean level of parents’ education reported by participants (1 = less than high school; 2= high school degree or GED ; 3=associate’s, technical, vocational, or trade school degree; 4= bachelor’s degree; 5= graduate degree). ^c Frequency of sexual activity refers to the number of times in the past three months a person reports having any type of sex ^d Inconsistency of condom use refers to the proportion of times in the past three months a person reports having any type of sex without using a condom. ^e Regression adjusted means are reported; standard deviations are not adjusted. ^f Regression adjusted mean differences are reported; rounding accounts for slight discrepancies in reported differences.

G. Methods

1. Impact evaluation

The impact study investigates whether or not offering BART impacts participants’ reported inconsistency of condom use (primary research question) and frequency of sex (secondary research question). We do this within an 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 BART relative to the offer of receiving the control condition Healthy Living. To answer both

primary and secondary research questions, we use a regression-estimated approach that models outcomes as a function of the baseline measure of the outcome variable (i.e., inconsistency of condom use and frequency of sex at baseline) as well as the following individual-level covariates measured at baseline and blocking variables: age, sex, race, Hispanic, parents' education, family structure, employment site, employed, cohort, and work shift (see the *Model Specification* section of Appendix D for details on variable construction). Since assignment is randomized, a simple difference of means of the outcome variables should provide an unbiased estimate of program impact; however, we statistically adjust for covariates to increase the precision of our estimates and to account for blocking procedures. See Appendix D for details of our analytic approach, including model specifications.

Assuming that assignment procedures are conducted with fidelity, missing data pose the greatest threat to the internal validity of an RCT within an ITT framework. Therefore, as detailed in the *Missing Data Approach* section of Appendix D, we mitigate the loss of cases due to item non-response with dummy variable adjustment for missing pretest and covariate data and multiple imputation for missing outcome data.

To test the robustness of our analytic approach, and to add certainty to our findings, we conducted several sensitivity analyses that test whether or not our findings are sensitive to our decisions; details of these analyses along with results can be found in Appendix E.

2. Implementation evaluation

The implementation study provides important context for the impact findings. Our analytic approach is to provide a descriptive analysis of the extent to which the program was implemented as intended. We present counts, calculate proportions and means, and provide written descriptions based on document review. Descriptive statistics are reported overall and by cohort. There are several limitations of the implementation data: 1) health educator self-reports

may not be a reliable measure of the content that was actually delivered to participants; additionally, we do not have complete self-report data for all BART and Healthy Living intervention sessions delivered; 2) observer data are very incomplete and may thus fail to offer a representative picture of the content actually delivered to youth; we have limited observation data for all BART and Healthy Living session types; and 3) data that are used to assess quality of staff-participant interactions are based on a partial convenience sample and may not be representative of all interactions. See Table F.1 in Appendix F for detailed methods for each implementation evaluation element.

IV. Study findings

A. Implementation study findings

Below, we present an overview of findings from our implementation study (detailed results are included in Appendix G).

Adherence

Sessions offered. Data presented in Table G.1 show that nearly all (99%) of the 344 intended BART sessions were offered. All eight sessions were offered to 41 of the 43 classes receiving BART; two classes were not offered session 7, and two were not offered session 8 due to low attendance or because the classroom was unexpectedly unavailable (see the *External Events Affecting Implementation* section in Appendix G for more details). Though the program was intended to be offered in eight weeks, the duration of programming for this study was five to eight weeks (multiple sessions were offered in select weeks during condensed programming). A plurality of BART classes were offered the eight-session program over six weeks (49%), followed by smaller proportions that were offered the program over five (14%), seven (16%), and eight (21%) weeks (see Table G.2 in Appendix G).

Sessions received. On average, participants assigned to BART received between six and seven (mean = 6.3) of the eight intended programming sessions (see Table G.5 in Appendix G). Eighty-five percent of participants attended session 1 which provided information on HIV and was identical to Healthy Living session 1. Only 2.8% of the treatment group attended no BART programming, and 40% attended all sessions (see Table G.6 in Appendix G).

Content delivered. There is variation in the proportion of activities completed across sessions. The average percent of activities completed per session range from 75% in session 1 (on average 4.5 of 6 activities were completed) to 98% in session 4 (on average 3.9 of 4 activities completed) (see Table G.7 and G.9 in Appendix G.). Similarly, session 1 was the least often completed (all activities were delivered in only 42% of cases), and session 4 was the most often completed (all activities were delivered in 93% of cases) (see Table G.9 in Appendix G).

Program staff. A total of 41 health educators facilitated the interventions. All health educators were trained in both the BART and Healthy Living curricula; 88% of health educators also completed fidelity monitoring and evaluation research basics trainings (see Tables G.10 and G.11 in Appendix G).

Quality of implementation

Approximately 23% of all BART sessions were assessed for quality using items from the *Program Observation Form for TPP Grantees*. For all quality measures, response options range from 1 (worst rating) to 5 (best rating); all four indicators were calculated as the percentage of observed sessions the fidelity monitor scored as a 4 or 5. Of the 80 BART sessions observed, 64% were scored as good or very good for the *delivery of session information*; *extent of participants' understanding* was scored as moderate or good in 70% of the assessed sessions; *extent of group members' participation* was scored as moderate or active for 65% of the assessed

sessions; and *overall quality of the program session* was scored as good or excellent for 65% of the assessed sessions (see Table G.12 in Appendix G).

Experiences of counterfactual group

In all, 98% of the 344 intended Healthy Living sessions were offered (see Table G.13 in Appendix G). Eighty-three percent of participants attended session 1 which provided information on HIV and was identical to BART session 1. Eighty-four percent of participants attended session 2; attendance decreased slightly at each subsequent session to 69% at session 8. On average, Healthy Living participants received six to seven (mean = 6.2) of the eight possible sessions; 3.8% of participants attended no Healthy Living sessions whereas 36% attended all sessions (see Tables G.16 – G.18 in Appendix G).

Content delivered. Similar to the delivery of session 1 in BART, on average 75% of the six prescribed session 1 activities were delivered; all six activities were completed in approximately 45% of sessions. Though not part of the Healthy Living curriculum and not instructed to do so, health educators discussed core elements of the BART program, to a limited extent, outside of session 1. Health educators discussed HIV/AIDS knowledge in 17.5% of session 2s, 2.4% of session 4s, and 2.4% of session 5s assessed. Health educators also discussed other core components, namely negotiation and condom use skills, in 5.0% of session 2s and 2.3% of session 3s assessed (see Table G.21 in Appendix G.)

Quality of implementation was assessed for 22% of all Healthy Living sessions using items from the *Program Observation Form for TPP Grantees*. Of the 75 Healthy Living sessions observed, 71% were scored as good or very good for the *delivery of session information*; *extent of participants' understanding* was scored as moderate or good in 59% of the assessed sessions; *extent of group members' participation* was scored as moderate or active for 75% of the assessed

sessions; and *overall quality of the program session* was scored as good or excellent for 60% of the assessed sessions (see Table G.12 in Appendix G).

Context

Overall, a majority of participants reported recent exposure to formal reproductive health education at each data collection point of interest: 56% at baseline (53% BART and 58% Healthy Living), 61% at post-program (62% BART and 60% Healthy Living), and 67% at six-month follow-up (73% BART and 61% Healthy Living) (see Table G.23 in Appendix G). A small proportion of participants reported exposure to specific TPP programming. The percent of participants reporting past year exposure to at least one TPP program (other than BART) at each data collection point is as follows: 15% at baseline (16% BART and 13% Healthy Living), 17% at post-program (15% BART and 19% Healthy Living), and 10% at six-month follow-up (10% BART and 10% Healthy Living) (see Table G.24 in Appendix G).

In September 2012, LPHI requested and was approved by OAH to implement an adaptation to the BART program (and, correspondingly, to the comparison intervention). This adaptation reduced the program duration from eight weeks, as was originally intended, to six weeks; it did not alter the BART curriculum content or reduce the number of sessions offered. This adaptation was requested to help ensure that youth had the opportunity to receive all eight program sessions during the summer period. Historically, the standard length of the NYW summer employment program was six weeks; however, in the first year (summer 2012), LPHI provided funding to extend the program at HEP sites for two additional weeks, and BART was implemented over eight weeks with one session per week, as prescribed. When planning for Year Two, LPHI found that the typical school year did not leave enough time during summer break for an adequate number of teens to maintain participation in an eight-week program; therefore, in the second and

third years (summers 2013 and 2014), at most sites, eight sessions were implemented over the course of six weeks, such that during two program weeks, two sessions were offered instead of one session. The primary concern with this adaptation is that for program weeks in which two sessions were offered instead of one, participants had less time to process session content between sessions. In addition, due to unforeseen circumstances, during summer 2013, programming was shortened from six to five weeks at two sites because classroom space was lost midway through implementation. As a result, at these sites sessions seven and eight were either combined, provided to multiple classes in a large group setting, or not provided at all. (Further details on external events affecting implementation and adaptations are provided in the *Context* section of Appendix G.)

B. Impact study findings

Inconsistency of Condom Use

Findings suggest that the offer to participate in BART had no significant effect on participants' inconsistency of condom use at six-month follow-up. Estimates presented in Table IV.1 demonstrate statistically insignificant differences in the proportion of times treatment and control participants report having sex without condoms in the past three months. Regression adjusted means for the treatment and control group of 0.09 and 0.07, respectively, indicate participants were not regularly engaging in the risk behavior six months postprogram (i.e., they were either using condoms with relative consistency or were not engaging in sex), and the mean difference between groups (.02) is small and statistically insignificant ($p > .05$). Sensitivity analyses, presented in Appendix E, corroborate this finding and indicate that results are not sensitive to analytical decisions. In each of the sensitivity studies, the mean difference in

participants’ inconsistency of condom use reported by treatment and comparison groups remains statistically insignificant.

Frequency of Sexual Activity

Findings also indicate that the offer to participate in BART had no impact on participants’ frequency of sex at six-month follow-up. Estimates presented in Table IV.2, demonstrate no statistically significant difference in the number of times treatment and control participants report having sex in the past three months. Regression adjusted means indicate that, on average (taking into account covariates and blocking variables), participants in both groups had sex less than two times in the previous six months, and the difference between groups (-0.2) is small and not statistically significant ($p > .05$). Sensitivity analyses, presented in Appendix E, again confirm this finding.

Table IV.1. Post-intervention estimated effects using data from the 6-month follow-up HEP Questionnaire to address the primary research questions

Outcome measure	Intervention mean or % (standard deviation)	Comparison mean or % (standard deviation)	Intervention compared to comparison mean difference (p -value of difference)
Inconsistency of condom use	0.09(0.26)	0.07(0.22)	0.02(0.202)
Sample Size	342	346	

Source: Follow-up surveys administered 6 to 12 months after the program.

Notes: The outcome measure is a risk variable – the higher the mean proportion, the more inconsistently participants are engaging in protected sex. The means reported in the table represent the regression adjusted means of the outcome variable; the standard deviations represent the unadjusted pooled standard deviation of the outcome variable (this is calculated from the 10 individual imputations used in our multiple imputation (benchmark) analysis). See Table III.1. for a more detailed description of our outcome measure and Appendix D for details of our analytic methods, including our missing data approach.

Table IV.2. Post-intervention estimated effects using data from the 6-month follow-up HEP Questionnaire to address the primary research questions

Outcome measure	Intervention mean or % (standard deviation)	Comparison mean or % (standard deviation)	Intervention compared with comparison Mean difference (p-value of difference)
Frequency of sexual activity	1.25(6.02)	1.41(5.04)	-0.17(0.653)
Sample Size	342	346	

Source: Follow-up surveys administered 6 to 12 months after the program.

Notes: The means reported in the table represent the regression adjusted means of the outcome variable; the standard deviations reported in the table represent the unadjusted pooled standard deviation of the outcome variable (this is calculated from the 10 individual imputations used in our multiple imputation (benchmark) analysis). See Table III.2 for a more detailed description of our outcome measure and Appendix D for details of our analytic methods, including our missing data approach.

V. Conclusion

Findings from this study indicate that the offer to participate in BART did not have a significant impact on the sexual behaviors of youth. Six months following the conclusion of the program, there was no statistically significant difference between treatment and comparison group members with regard to their self-reported inconsistency of condom use or frequency of sex. Sensitivity analyses all corroborate these findings.

Based on a previous study of BART (cited by the HHS sponsored evidence review), we expected that following programming members of the treatment group would exhibit less risky behavior (i.e., decreased frequency of sex and less inconsistent condom use) than members of the control group, who were offered a health intervention with the same initial HIV informational component as BART but none of the additional motivational aspects of the program.ⁱⁱⁱ However, our results fail to replicate the previous evidence that found the program to be effective at promoting safe sex behaviors.ⁱⁱⁱ

We do not have a convincing explanation for the divergent results. Implementation results reported here indicate the program in our study was conducted with reasonable fidelity.

Furthermore, exploratory analysis demonstrates that the intervention had significant effects on

many of the hypothesized antecedents of behavior change.^{xiii} Yet, findings consistently demonstrate that, six months after intervention, the anticipated reductions in sexual behaviors for the treatment group do not occur. Treatment and control groups report statistically insignificant differences in condom use and frequency of sex. In the remainder of this section, we present potential explanations for the divergent results and suggestions for further research that can address new questions generated by this study.

Given that the outcome distributions are skewed to the right and the mass concentrated towards zero, we conducted several additional analyses to assess whether our benchmark analysis might explain the discrepancy in findings. First, although Lumley et al. (2002) emphasize that the approach is unnecessary and Schneider et al. (2007) advise against it, we conducted our benchmark analyses using log transformations of our two outcome variables (using the formula $\log_{10}(x+1)$) to determine if this explains the variation in results.^{xiv xv} Next, we specified robust standard errors in our benchmark model. Third, we fitted a poisson regression with the benchmark model. Fourth, we conducted non-parametric tests of group equivalence (Wilcoxon rank-sum test) on our non-transformed outcome variables. Results from these tests corroborate our benchmark findings; in each case there is no statistically significant difference in outcomes for the treatment and control groups.

We considered the fact that the sample for the previous study was comprised primarily of females and the possibility that this study's effects at six months were contingent on gender. However, an exploratory analysis reveals no statistically significant difference in outcomes between males and females in the treatment group.

Though the authors of the prior study do not report average program dosage, we also considered whether or not dosage could explain our results. Although the sub-groups are

endogenous, for exploratory purposes we tested whether or not the number of sessions attended and full program exposure affected outcomes in our study and found that dosage is not significantly related to the outcomes assessed, and there is no interaction between dosage and treatment.

Although the differences are slight, the two samples are dissimilar in a way that may be consequential. Youth in our study are slightly younger and report less sexual experience than the youth in the study that found significant impacts. The authors of the prior study report that the mean age is 15.3 years, and the mean number of lifetime sexual partners reported by their sample is 2.7; by contrast, in this study baseline mean age is 15.0 years, and the mean number of lifetime partners is 2.0. This could matter because youth in our study with (reportedly) less sexual experience may not report enough sex for an impact to be statistically manifest. But, an analysis of the empirical data does not support this line of thought. When we remove the 380 individuals who report that they were not sexually active at baseline and conduct the benchmark analysis ($n = 308$), the coefficient for the treatment indicator remains statistically insignificant; similarly, in the full sample ($n = 688$) the treatment effect is not conditional upon sexual experience at baseline (i.e., there is no interaction between treatment and sexual experience).

There are also plausible explanations that lie beyond the scope of our data. The sample for the original causal study was drawn from a health clinic; youth were recruited into the study because they were receiving health care at the clinic. It is possible that, because youth were receiving the intervention in this setting, they were more receptive to BART's health promotion messages as compared to our sample who were receiving these messages in the context of a summer employment program. This explanation is somewhat supported by the fact that modified

versions of the program have also demonstrated some evidence of effectiveness in adolescent drug dependency/substance use treatment programs.^{xvi xvii}

To the extent that they are of equal quality, study results such as this that fail to reject the null hypothesis and find that the intervention has failed to effect the behavioral change hypothesized should be of equal evidentiary value to those that find otherwise. They may, in fact, provide more opportunity or incentive to learn why the intervention works in some cases and not in others and what conditions are necessary for causal impacts. Future studies conducted under the auspices of this grant will examine whether there were any long-term impacts on inconsistency of condom use and frequency of sex (12 months postprogram). We can hypothesize that the impacts of the program may become more evident as youth grow and become more sexually active. Results from the previous causal study on BART indicate this potential. Another avenue for exploratory study is to investigate the predictive or associational relationships between the theoretically relevant behavioral antecedents (attitudes, beliefs, intentions, and self-efficacies) and the behavioral outcomes of interest. This may help us better understand the results reported here. In addition, given the extent of sexual inexperience in our sample, we may examine whether the program effectively promotes delayed initiation of sex for this subpopulation.

VI. References

- ⁱ US Department of Health and Human Services, Office of Adolescent Health . TPP Resource Center: Evidence-Based Programs. Available at: http://www.hhs.gov/ash/oah/oah-initiatives/teen_pregnancy/db/index.html. Accessed March 20, 2015.
- ⁱⁱ Goesling B, Colman S, Trenholm C, Terzian M, Moore K. Programs to Reduce Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors: A Systematic Review. *J Adolesc Health*. 2012; 54:499-507.
- ⁱⁱⁱ St Lawrence JS, Brasfield TL, Jefferson KW, et al. Cognitive-behavioral intervention to reduce African-American adolescents' risk for HIV infection. *J Consult Clin Psychol*. 1995; (63):221-237.
- ^{iv} Louisiana Department of Health and Hospitals. 2014 Health Report Card. Available at: <http://www.dhh.state.la.us/index.cfm/page/2150>. Accessed June 5, 2015.
- ^v Louisiana Department of Health and Hospitals, Office of Public Health STD/HIV Program. 2013 Annual Report: Sexually Transmitted Diseases. Available at: http://www.dhh.state.la.us/assets/oph/Center-PHI/2014HealthReportCard/2013_Louisiana_STD_Annual_Report.pdf. Accessed June 5, 2015.
- ^{vi} Ajzen I, Joyce N, Sheikh S, Cote N. Knowledge and the prediction of behavior: the role of information accuracy in the theory of planned behavior. *Basic Appl Soc Psych*. 2011; 33(2):101-117.
- ^{vii} Bandura A. Self-efficacy mechanism in human agency. *Am Psychol*. 1982; 37(2):122-147.
- ^{viii} Bandura A, Adams NE, Beyer J. Cognitive processes mediating behavioral change. *J Pers Soc Psychol*. 1977; 35(3):125-139.
- ^{ix} Fisher W, Williams S, Fisher J, Malloy T. Understanding AIDS risk behavior among sexually active urban adolescents: an empirical test of the information-motivation-behavioral skills model. *AIDS Behav*. 1999; 3(1):13-23.
- ^x Taylor D, Bury M, Campling N, et al. A review of the use of the Health Belief Model (HBM), the Theory of Reasoned Action (TRA), the Theory of Planned Behavior (TPB) and the Trans-Theoretical Model (TTM) to study and predict health related behavior change. National Institute for Health and Clinical Excellence. *NICE guidelines PH6*. 2006; 1-215.
- ^{xi} Scott CK. A replicable model for achieving over 90% follow-up rates in longitudinal studies of substance abusers. *Drug Alcohol Depend*. 2004; 74(1):21-36.
- ^{xii} Davis E, Demby H, Jenner LW, Gregory A, Broussard M. Adapting an evidence-based model to retain adolescent study participants in longitudinal research. *Eval Program Plann*. Forthcoming 2016.

^{xiii} Walsh S, Jenner E, Leger R, Broussard M. Effects of a Sexual Risk Reduction Program for African-American Adolescents on Social Cognitive Antecedents of Behavior Change. *Am J Health Behav.* 2015; 39(5):610-22. doi: 10.5993/AJHB.39.5.3.

^{xiv} Lumley T, Diehr P, Emerson E, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health.* 2002 ;23:151-69.

^{xv} Schneider D, Tahk A, Krosnick Reconsidering the impact of behavior prediction questions on illegal drug use: The importance of using proper analytic methods. *Social Influence.* 2007; 2(3):178 – 196

^{xvi} St. Lawrence JS, Jefferson KW, Alleyne E, Brasfield TL. Comparison of education versus behavioral skills training interventions in lowering sexual HIV-risk behavior of substance-dependent adolescents. *J Consult Clin Psychol.* 1995; 63(1):154-157.

^{xvii} St Lawrence J, Crosby R, Brasfield T, O'Bannon III R. Reducing STD and HIV risk behavior of substance dependent adolescents: a randomized controlled trial. *J Consult Clin Psychol.* 2002; 70(4):1010-1021.

Appendix A: Data collection efforts

Impact Study Data Sources

Participant data were collected via a self-administered questionnaire or a brief phone interview. The questionnaire comprises 116 items that ask participants to report on various demographic characteristics, sexual behaviors, and theoretical antecedents to those behaviors. Prior to administration, the questionnaire was field-tested with 10 health professionals (including MDs, MPHs, PhDs) as well as 12 adolescents (six boys and six girls, ages 14 to 15) to ensure the questions were valid, relevant, and comprehensible by youth. Though there were slight modifications made to the questionnaire during the study period (between administrations questions were reordered, and a number of questions not essential to the study were removed), no substantive changes were made.

The phone interview was an abbreviated version the questionnaire (25 items) – it contained those questions necessary for our impact analysis as well as select questions gauging participants’ perceptions and attitudes associated with safe sex practices.

Table A.1. Data collection procedures used in the impact analysis of BART

	Baseline	6-month follow-up
Maximum data collection window	5-8 weeks depending on length of programming	6 months
Survey opens	Day of or before first program session (before session).	6-months following date of last program session
In person, self-administered paper questionnaire opens	Survey open date	Survey open date
Self-administered web-based questionnaire opens	As needed if participant could not come in to complete	2 months following open date
Mail-in, self-administered paper questionnaire opens	n/a (survey mode not offered)	4 months following open date
Brief phone interview opens	n/a (survey mode not offered)	5 months following open date
Survey closes	Day of last program session (before session).	6-months following open date
Follow-up methods of contact	Phone call, text, email, mail	Phone call, text, email, mail
Incentives for completed questionnaires	Entry into raffle for iPod Touch	Entry into raffle for iPod Touch \$20 Walmart gift card
Differences in procedures between treatment and control	None	None

Note: In order to decrease study attrition participants who were randomized but who did not attend any programming received an additional \$20 Walmart gift card for completing their baseline questionnaire. In addition, participants who had not responded to the 6-month follow-up questionnaire close to the end of their follow-up window (1 month prior to their scheduled close date) were offered an additional \$15 movie theater gift card incentive.

Table A.2. Data collection efforts used in the impact analysis of BART six months postprogram

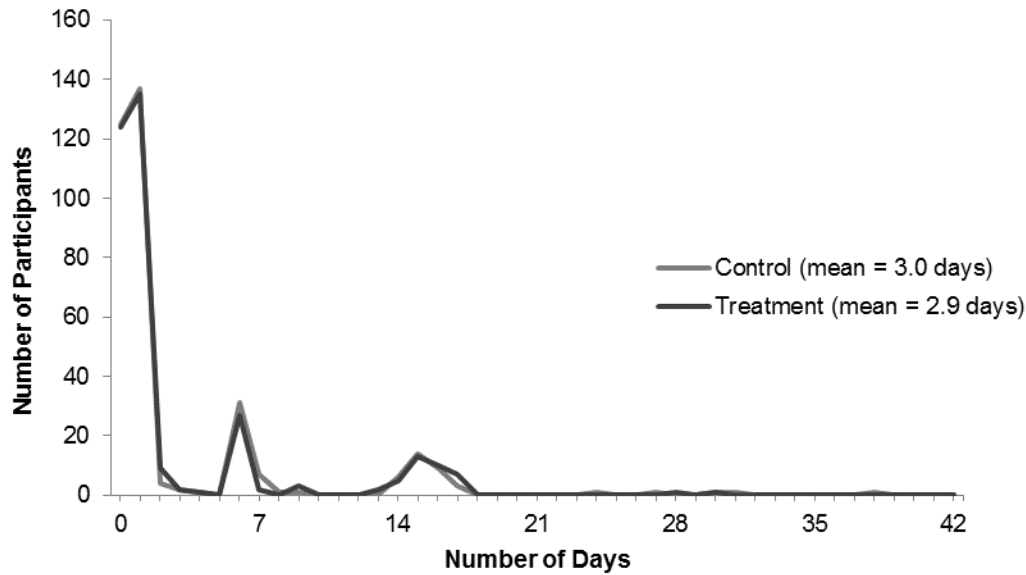
Data collection effort	Cohort 1 (Work and Learn)	Cohort 2 (Teen Camp)	Cohort 2 (Work and Learn)	Cohort 3 (Teen Camp)	Cohort 3 (Work and Learn)
Start date of programming	6/11/2012	6/5/2013	6/24/2013	6/3/2014	6/16/2014
Baseline survey	6/11– 7/9/2012	6/4– 7/5/2013	6/24– 7/24/2013	6/2– 6/26/2014	6/16– 7/20/2014
Six-month follow-up survey	2/16– 7/12/2013	1/12– 6/26/2014	2/1– 8/1/2014	1/10– 4/23/2015	1/23– 4/23/2015

Note: LPHI partnered with NOLA Youth Works (NYW) and several of their CBOs to implement HEP (both BART and Healthy Living) as a component of their summer programming. All youth applied to the NYW program in the same way, and NYW placed (assigned) youth to participating CBO sites based on their age, residence, program preference, and site capacity; some HEP sites provided ‘teen camp’ programming, whereas other sites provided ‘work and learn’ programming. In 2012 (Cohort 1), HEP participants were placed only at sites offering ‘work and learn’ programming; in 2013 (Cohort 2) and 2014 (Cohort 3), HEP participants were placed at both ‘teen camp’ and ‘work and learn’ programming sites. Teen camp sites began summer programming two to three weeks earlier than work and learn sites, which is why data collection efforts are separated for these two cohorts. Dates provided for the baseline and six-month follow-up survey administration reflect the dates the first and last participant from each group submitted questionnaires.

Table A.3. Questionnaire completion by mode of administration for data collected for impact analysis of BART

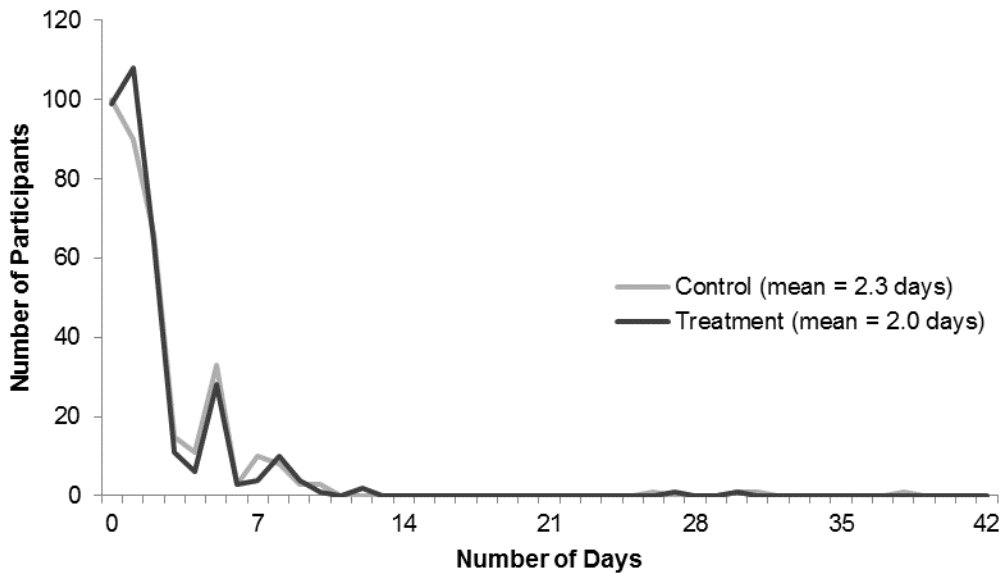
Mode of administration	Number in Control Group, Baseline (% of group)	Number in Treatment Group, Baseline (% of group)	Number in Control Group, Six-Month Follow-up (% of group)	Number in Treatment Group, Six-Month Follow-up (% of group)
In-person	344 (99.4%)	341 (99.7%)	234 (67.6%)	245 (71.6%)
Online	2 (0.6%)	1 (0.3%)	98 (28.3%)	90 (26.3%)
Mail	0 (0.0%)	0 (0.0%)	11 (3.2%)	6 (1.8%)
Phone Interview	not applicable	not applicable	3 (0.9%)	1 (0.3%)
Sample size	346	342	346	342

Figure A.1. Number of days from randomization to baseline questionnaire completion, by study condition



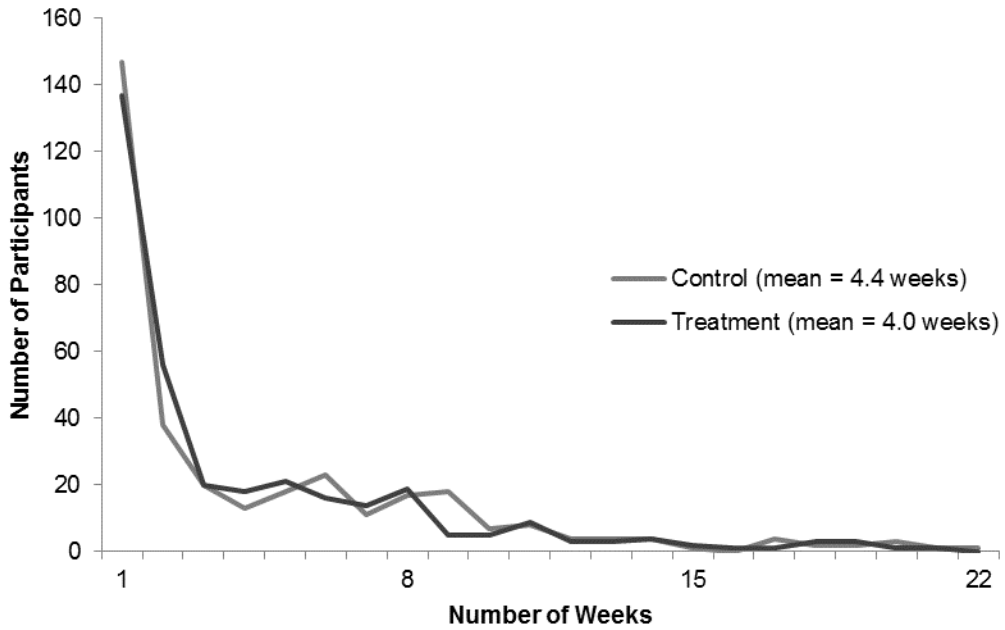
Note: To assess whether there were statistically significant differences between groups, we performed an OLS regression in which we regressed the number of days from randomization to baseline completion on the treatment indicator as well as blocking covariates (cohort, site, shift). The regression adjusted mean difference between the treatment and comparison groups (-0.25) is small and statistically insignificant ($p > .05$).

Figure A.2. Number of days from the open of baseline survey window to questionnaire completion, by study condition



Note: To assess whether there were statistically significant differences between groups, we performed an OLS regression in which we regressed the number of days from the open of the baseline survey window to the completion of the questionnaire on the treatment indicator as well as blocking covariates (cohort, site, shift). The regression adjusted mean difference between the treatment and comparison groups (-0.31) is small and statistically insignificant ($p > .05$).

Figure A.3. Number of weeks from the open of six-month survey window to questionnaire completion, by study condition



Note: To assess whether there were statistically significant differences between groups, we performed an OLS regression in which we regressed the number of weeks from the open of the six-month survey window to the completion of the questionnaire on the treatment indicator as well as blocking covariates (cohort, site, shift). The regression adjusted mean difference between the treatment and comparison groups (-0.27) is small and statistically insignificant ($p > .05$).

Appendix B: Implementation evaluation data collection

Implementation Study Data Sources

Class Information Form. The Class Information Form was used to collect data about the makeup of each class, including: names of each participant assigned to the class, intervention assignment (BART/ Healthy Living), health educators IDs, assigned time of class, class size, site ID number, and class ID number. This form was completed once for each class after randomization.

Attendance Sheet. The Attendance Sheet for each class was used to collect the following administrative data for each session: session date, facilitator names and IDs, site name, class ID, class gender, and the number of the sessions completed (1-8), the individual participants in attendance at each session, and the total number of participants present at each session. Health educators were required to record these data for each session offered.

BART Implementation Fidelity Tool. The fidelity tool was used to collect the following fidelity data for each session (eight sessions total): session date, facilitator names, site name, class ID, class gender, number of students in class, and, for each session activity: activity completed, activity completed with changes, or activity not completed. There were separate forms for health educators and fidelity monitors. Health educators were to complete 1 form for each session completed and fidelity monitors were to complete forms for 20% of all sessions.

The Healthy Living Implementation Fidelity Tool. The fidelity tool was used to collect the following fidelity data for each control session: session date, facilitator names, site name, class ID, class gender, number of students, fidelity data for each activity implemented within session 1 (indicates if activity completed, completed with changes, or not completed), and, for sessions 2 through 8 data were collected to indicate if the facilitator engaged in any of BART's core components during that session (yes or no). There were separate forms for health educators and fidelity monitors. Health educators were to complete 1 form for each session completed and fidelity monitors were to complete forms for 20% of all sessions.

Program Observation Form for TPP Grantees. This form was used to collect data on the overall quality of the program session and delivery of the information. These were to be completed for 20% of all BART and Healthy Living sessions completed by a fidelity monitor.

Table B.1. Data used to address implementation research questions

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Adherence: How often were sessions offered? How many were offered?	<i>Session date and the number of session offered (1-8) from the Attendance Sheet for each class.</i>	<i>Every session offered.</i>	<i>Program staff (Health Educator teams of two) who offer the session.</i>
Adherence: What and how much was received?	<i>Session date, number of the session completed (1-8), and the individual participants in attendance at each session from the Attendance Sheet for each class</i>	<i>Every session that is offered.</i>	<i>Program staff (Health Educators) who offer the session.</i>
Adherence: What content was delivered to youth?	<i>Completion status for each session activity: activity completed, activity completed with changes, or activity not completed - from the BART Implementation Fidelity Tool: Health Educator Self-Report</i>	<i>Recorded after every session.</i>	<i>Program staff (Health Educators) who offer the session.</i>
	<i>Completion status for each session activity: activity completed, activity completed with changes, or activity not completed – from the BART Implementation Fidelity Tool: Observer Report</i>	<i>Recorded by fidelity monitors for a convenience sample of at least 20% of BART sessions conducted within each implementation cohort.</i>	
Adherence: Who delivered material to youth?	<i>Lists of Health Educators hired to implement program for each cohort, including their credentials (degree/certifications).</i>	<i>Data are available to LPHI program staff.</i>	<i>Program staff (fidelity monitors) following each session that is observed.</i>
	<i>List of Health Educator position qualification requirements (as created by program staff).</i>	<i>Determined prior to hire date and are available to LPHI program staff.</i>	
	<i>Lists of Health Educator staff from each cohort who have completed the following trainings: BART curriculum, Healthy Living curriculum, Fidelity Monitoring, and Evaluation Research Basics.</i>	<i>Training attendance data are available to program staff.</i>	
Quality: Quality of staff-participant interactions	<i>Quality of staff-participant interactions data from questions 1-5 and 7 on the Program Observation Form for TPP Grantees (developed by OAH).</i>	<i>Recorded for each classroom session selected for observation. (Convenience sample of 20% of BART and Healthy Living sessions selected for observation for each cohort.)</i>	<i>Program staff (Fidelity Monitors) following each session that is observed.</i>
Counterfactual: How often were sessions offered? How many were offered?	<i>Session date and the number of session offered (1-8) from the Attendance Sheet for each class.</i>	<i>Every session offered.</i>	<i>Program staff (Health Educator teams of two) who offer the session.</i>

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Counterfactual: What and how much was received?	<i>Session date, number of the session completed (1-8), and the individual participants in attendance at each session from the Attendance Sheet for each class.</i>	<i>Every session that is offered.</i>	<i>Program staff (Health Educators) who offer the session.</i>
Counterfactual: What content was delivered to youth? (Note: The Healthy Living fidelity tool collects fidelity data on all activities completed in session 1 (which is exactly the same as BART session 1); we do not monitor fidelity to the Healthy Living curriculum for sessions 2-8.)	<i>Activity completion status for each activity implemented within session 1 (activity completed, completed with changes, or not completed), and for sessions 2-8 data are collected to indicate if the facilitator engaged in any of BART's core components during that session (yes or no) – collected with the Healthy Living Implementation Fidelity Tool: Health Educator Self-Report.</i>	<i>Data are to be recorded after every session.</i>	<i>Program staff (Health Educators) following each session that is delivered.</i>
	<i>Activity completion status for each activity implemented within session 1, and for sessions 2-8 data are collected to indicate if the facilitator engaged in any of BART's core components during that session (yes or no) – collected with the Healthy Living Implementation Fidelity Tool: Observer Report</i>	<i>Data are to be recorded by fidelity monitors for a convenience sample of at least 20% of Healthy Living sessions conducted within each implementation cohort.</i>	<i>Program staff (fidelity monitors) following each session that is observed.</i>
Counterfactual: Who delivered material to youth?	<i>Lists of Health Educators hired to implement program for each cohort, including their credentials (degree/certifications).</i>	<i>Data are available to LPHI program staff</i>	
	<i>List of Health Educator position qualification requirements (as created by program staff).</i>	<i>Determined prior to hire date and are available to LPHI program staff.</i>	
	<i>Lists of Health Educator staff from each cohort who have completed the following trainings: BART curriculum, Healthy Living curriculum, Fidelity Monitoring, Evaluation Research Basics.</i>	<i>Training attendance data are available to program staff.</i>	
Context: Other TPP programming available or offered to study participants (both intervention and comparison)	<i>List of other TPP programming being implemented in Orleans Parish during program period.</i>	<i>List developed during grant Year 1 and updated on an ongoing basis.</i>	<i>Recorded by Evaluation staff (PRG Research Analyst).</i>
	<i>Two items on the Health Education Program Questionnaire collect individual-level self-reported data on participants' reproductive health education and experiences with other TPP programs in the past year.</i>	<i>Data collected from participants at baseline, postprogram, 6 months postprogram and 12 months postprogram.</i>	<i>Recorded by Evaluation staff (PRG Research Assistants)</i>

Implementation element	Types of data used to assess whether the element of the intervention was implemented as intended	Frequency/sampling of data collection	Party responsible for data collection
Context: External events affecting implementation	<i>HEP Study Methods Log, OAH progress reports (6-month and annual), and HEP project meeting notes.</i>	<i>Ad hoc</i>	<i>Methods Log data recorded by Evaluation staff; progress reports recorded by LPHI program staff; project notes by both program and evaluation staff.</i>
Context: Substantial unplanned adaptation(s)	<i>Adaptation requests to OAH, OAH progress reports (6-month and annual), HEP project meeting notes.</i>	<i>Adaptation requests completed as needed; progress reports completed every six months; meeting notes taken at biweekly/weekly project meetings</i>	<i>Adaptation requests and progress reports recorded by LPHI program staff; meeting notes are recorded by both program and evaluation staff.</i>

TPP = Teen Pregnancy Prevention.

Appendix C: Study sample

Table C.1. Youth data collection by intervention status

Number of youth	Time Period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Assigned to condition		850	427	423			
Contributed a baseline survey		842	424	418	99.1%	99.3%	98.8%
Contributed a follow-up survey	6 months post-programming	742	370	372	87.3%	86.7%	87.9%

Table C.2 Youth sample sizes by intervention status

Number of youth	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Assigned to condition	850	427	423			
Contributed both baseline and 6-month questionnaires	739	368	371	86.9%	86.2%	87.7%
Contributed both baseline and 6-month questionnaires and data deemed reliable	688	342	346	80.9%	80.1%	81.8%

Appendix D: Implementation evaluation methods

Model specification

The empirical models for both research questions were estimated with an OLS regression (using Stata). We present the empirical model for our primary research question below; the model for our secondary research question is identical except that the outcome variable and baseline measure of the outcome variable is frequency of sex (continuous, range 0 to k).

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

where:

Y_{Post} – The outcome variable, the inconsistency of condom use (continuous proportion; range 0 to 1, where 0= has sex without condoms 0% of the time and 1= has sex without condoms 100% of the time) reported by participant i at the 6-month post intervention.

Y_{Pre} – The baseline measure of the outcome variable, the inconsistency of condom use (continuous proportion; range 0 to 1, where 0= has sex without condoms 0% of the time and 1= has sex without condoms 100% of the time) reported by participant at baseline; variable re-centered at the grand mean for analysis.

T – A dummy treatment indicator variable whose value equals 1 if the participant was 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 include:

- a) Age – self reported age at baseline (continuous; range 14-18); variable re-centered at the grand mean for analysis.
- b) Race – self-reported race of participant. A dummy variable (0= not black or African American; 1= black or African American); variable re-centered at the grand mean for analysis.
- c) Ethnicity – self-reported ethnicity of participant. A dummy variable (0= not Hispanic or Latino; 1=Hispanic or Latino); variable re-centered at the grand mean for analysis.
- d) Parental education – A continuous measure of the mean level of parents’ education reported by participants (scores range from 1 = less than high school to 5 = graduate degree); variable re-centered at the grand mean for analysis.
- e) Family structure – A dummy indicator variable that measures whether a respondent lives with both parents (0= does not live with both parents; 1= lives with both parents); variable re-centered at the grand mean for analysis.
- f) Cohort – A set of 3-1 = 2 dummy blocking variables to capture the variable effects associated with the 3 cohorts exposed to the intervention during the evaluation period. Each dummy coded 1 if the individual was in the given cohort and coded 0 otherwise. Each of the dummy variables mean centered for analysis.
- g) Site – a set of 18-1 = 17 dummy blocking variables to capture the variable effects of the 18 sites that offered the interventions during the evaluation period. For each

variable, an individual participant was coded as 1 if s/he was assigned to particular site and 0 otherwise. Dummy variables grand mean centered for analysis.

- h) Shift – a dummy blocking variable to capture the variable effects associated with the assignment to morning or afternoon shift (0 = assigned to afternoon shift; 1= assigned to morning shift) ; variable re-centered at the grand mean for analysis
- i) Gender – a dummy blocking variable to capture the differential effects associated with participants’ gender (0=male; 1=female); variable re-centered at the grand mean for analysis.

β_0 – The intercept term, which represents the mean self-reported inconsistency of condom use for comparison participants, six months after the end of treatment, with all other variables in the model held constant at zero.

β_1 – This is the parameter estimate of substantive interest. β_1 represents the adjusted mean difference in treatment and comparison participants’ self-reported inconsistency of condom use six months after the end of treatment.

Data Cleaning Procedures

To improve the validity and reliability of our estimates, prior to analysis, we followed several steps to prepare the dataset and improve the quality of our data. First, we performed data quality checks, comparing recorded data against paper questionnaire entries to ensure no data entry errors were made. Next, we systematically screened or reviewed the analytic variables (outcome, baseline, or covariate) to identify invalid entries, inconsistencies, and unreliable data. These procedures are outlined below.

Identify and flag unreliable cases. The first step in the data screening process was to identify and flag cases (i.e., units or entire questionnaires) that were unreliable. By unreliable, we mean that we have sufficient reason to believe that the respondent’s answers were not honest representations of their behaviors, knowledge, and beliefs. Cases were flagged as unreliable in three instances: responses followed a clear, deliberate pattern; respondents finished the questionnaire in a time considered too fast to have read the questions and provided reliable responses (7 minutes for online questionnaires; 10 minutes for paper questionnaires); or the respondents indicated on their questionnaires that they were not honest as they responded. Data for cases that are deemed unreliable were treated as *unit missing* and excluded from benchmark analyses. However, sensitivity analyses that included the unreliable data were conducted and results are presented in Appendix E.

Identify and flag invalid responses. The second step in the data screening process was to inspect the data for instances in which responses were invalid because they were outside of a pre-determined range of plausible or acceptable values. Referring to a codebook containing variable names, valid variable values or ranges of values, and when applicable value labels, a research analyst performed diagnostics in Stata to ensure that responses to all analytic measures were valid. The analyst flagged all values that were out of range as invalid and recoded these values to missing. Data that were recoded to missing were treated according to our missing data approach.

Identify and flag outliers. The third step is to identify and flag severe outliers. By outliers, we are referring to values that are extreme compared to other observations, but are not invalid. Our benchmark analytic approach is to include data flagged as outliers (i.e., extreme values that are not considered invalid) in analysis, because we do not know for certain whether the values are true or invalid. However, we also ran sensitivity analyses that exclude these data and we report results in Appendix E.

Identify and flag inconsistencies in reporting of sexual behaviors. The final step in the data review process was to inspect the data and identify inconsistencies in sexual behavior outcome data. With repeated measures of sexual behaviors, two primary types of inconsistencies occur – internal inconsistencies and over-time inconsistencies. Internal inconsistencies refer to discrepancies in responses (to related questions) in the same survey administration. For instance, a respondent might say that s/he has not had sex in the past 3 months, but then indicates that s/he used condoms three of the times s/he had sex in the past 3 months. Over-time inconsistencies refer to instances in which a lifetime reported behaviors decline or are completely recanted over time. For example, at baseline a respondent might say that s/he has had sex 10 times in her/his life, but on the subsequent administration of the survey s/he says either a) s/he has never had sex or, b) s/he has sex 4 times in her/his life.

A research analyst examined outcome variables and flagged as inconsistent internally data in the following instances.

- If, on one questionnaire (baseline or six-month follow-up), a respondent indicates that s/he has had sex in the past 3 months (i.e., s/he provides a response greater than “0” to the question: *“In total, how many times have you had any type of sex in the past 3 months”*) but then indicates in the same survey administration that s/he has never had sex (i.e., s/he responds *“I have never had any type of sex”* to the question, *“How old were you the first time you had any type of sex?”*) all sexual behavior responses are flagged as inconsistent internally and recoded to missing.
- If, on one questionnaire (baseline or six-month follow-up), a respondent indicates that s/he has not had sex in the past 3 months (i.e., s/he responds “0” to the question, *“in total, how many times have you had any type of sex in the past 3 months?”*), but then indicates in the same survey administration that s/he has used condoms while having sex in the past three months (i.e., s/he provides a response greater than “0” to the question *“Now, think about the number of times that you had any type of sex in the past 3 months. How many of those times did you use condoms?”*), both responses are flagged as inconsistent internally and recoded to missing.
- If, on one questionnaire (baseline or six-month follow-up), a respondent indicates that s/he has used condoms more times in the past 3 months than she has had sex (i.e., his/her response to the question *“Now, think about the number of times that you had any type of sex in the past 3 months. How many of those times did you use condoms?”* is greater than the response given to the question *“in total, how many times have you had any type of sex in the past 3 months?”*) both responses are flagged as inconsistent internally and recoded to missing.

A research analyst examined outcome variables and flagged as inconsistent over time data in the following instances.

- If, at baseline, a respondent indicates that s/he has had sex during the past three months (i.e., s/he provides a response greater than “0” to the question, “*in total, how many times have you had any type of sex in the past 3 months?*”), then at six-month follow-up indicates that s/he has never had sex (i.e., s/he responds “*I have never had any type of sex*” to the question, “*How old were you the first time you had any type of sex?*”), outcome measures at both baseline and follow-up are flagged as inconsistent over time and recoded to missing.

Note on Data Recoding Error. During our data cleaning process, we discovered a data recording error that affected questionnaires administered online. In short, for our sexual behavior questions that asked respondents to indicate the number of times they had engaged in a particular behavior, our web survey software program recorded responses of 1 as missing responses. That is, if a person indicating having engaged in a behavior 1 time, no response was recorded for that individual. In these cases, we cannot determine whether individuals with missing data truly did not provide a response or whether they provided a response of 1 that was not recorded. For the purposes of our study, two questions on each the baseline and six month follow-up questionnaire were affected: “*In total, how many times have you had any type of sex in the past 3 months?*”; “*Now, think about the number of times that you had any type of sex in the past 3 months. How many of those times did you use condoms?*” In total, data for 27 respondents at baseline and/ or six-month follow-up were potentially affected (i.e., 27 individuals took the online questionnaire and have missing responses for at least one of the sexual behavior questions). Analyses presented elsewhere in this report treat problematic data as missing, and their values were imputed according to our missing data approach. However, to err on the side of caution, we also ran all benchmark analyses and sensitivity studies with these data recoded to 1. Results (not reported here) indicate that whether these cases are treated as missing or whether they are coded as 1, findings are substantively the same, and we conclude that our treatment of the problematic data did not affect our results.

Missing Data Approach

The benchmark approach to missing data that we selected aims to mitigate the introduction of bias into our impact estimates, provide good estimates of uncertainty, and maximize the use of available data by imputing or adjusting data. Our six-step decision processes is outlined below.

1. Using data cleaning procedures outlined in the *Data cleaning* section, identify inconsistent, unreliable, and invalid data in any analytic (i.e., outcome, pre-test, or covariate) variables and recode inconsistent and invalid data as missing and flag unreliable data for analysis.
2. Examine prevalence of unit and item missing (which result from nonresponse), as well as inconsistent, unreliable, and invalid data for both treatment and comparison samples.
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.
4. Determine whether any individuals who are in the randomized sample have no data at all (i.e., unit missing) at baseline and at the six-month follow-up observations. If this was the case, we reasoned that case-wise deletion is the most prudent approach, as no data exist (that are not imputed) at the individual-level from which to estimate values for the missing data.

5. For the remaining missing analytic data we then imputed or adjusted the missing values differently depending on whether the variables are: (a) pretest (and other covariate) data, or (b) post-test or outcome data.
 - a) For missing pre-test or covariate data, our benchmark approach is to use dummy variable adjustment procedures. Although Puma et al. concede that this approach is questioned in the literature, they recommend it as a preferred approach regardless of whether data are *missing at random*, *missing completely at random*, or *missing not at random*.¹ They argue and find in their simulations that it is an appropriate strategy to maximize the analytic sample without biasing results as long as the assignment to treatment is uncorrelated with the covariate missing data (which it should be, given that random assignment ensures that treatment is in expectation exogenous and unrelated to all observed covariates).
 - b) For missing post-test data, our benchmark approach is to use Multiple Stochastic Regression Imputation. Puma et al. recommend this as one approach that minimizes bias in their simulations. Briefly, this is a regression-based approach to imputation that imputes missing values with predicted values derived from the combination of multiple (in our case 10) iterations of the dataset (i.e., 10 separately constructed datasets with distinct predicted values). With this approach, variance is to be the same across imputed and observed values.
6. Conduct sensitivity analysis by estimating results with missing data excluded from the analysis (i.e., use case-wise deletion for all cases with missing data in analytic variables). In Appendix E, we report our benchmark results next to the sensitivity analysis results to verify findings.

References

1. Puma MJ. What to do when data are missing in group randomized controlled trials. *National Center for Education Evaluation and Regional Assistance: Institute of Education Sciences*. Washington, DC: US Department of Education; 2009:0049.

Appendix E: Sensitivity analyses

We conducted six sensitivity analyses to test the robustness and validity of our benchmark approach. Specifically we constructed alternative empirical models or altered data cleaning and imputation rules to examine the sensitivity of benchmark findings to the following analytic decisions: (1) the use of covariates to improve the precision of our estimates; (2) the use of imputation for missing data; (3) the use of unreliable data; and (4) the inclusion of outliers. In addition to these variations in specification, we also test the sensitivity of the benchmark results against those produced by: (5) programmatic variations in curriculum length of BART (from eight to six weeks) and (6) variations in the length of the data collection window. For the first four studies, we are interested in whether the results produced by alternative specifications produce different inferences than the benchmark results. If they do, we would conclude that the benchmark results are sensitive to our analytic decisions. For the last two sensitivity studies, we contrast the results of two subsamples rather than comparing them directly to the benchmark results. In these latter two studies, the presence of a statistically significant point estimate itself would indicate that the results are sensitive to variations in program duration (5) or response time (6).

Sensitivity Study 1: Baseline Covariates

We test our benchmark approach of including covariates (including the baseline measure of the outcome variable) in the analytic model by estimating an otherwise identical empirical model without the covariates included and comparing the sensitivity model estimates with the benchmark model estimates. Coefficients and p-values for the treatment indicators for the two contrasts are presented in tables E.1 and E.2 below under Sensitivity Study 1. The estimates produced by both models are substantively identical and indicate no programmatic effect on

inconsistency of condom use or frequency of sex six months after the program ends. The p -values in both the second and fourth columns are considerably greater than 0.05 for both outcomes. Consequently, we infer that substantive findings are identical regardless of whether or not we control for covariates in the analytic model.

Sensitivity Study 2: Missing Data

As detailed Appendix F, we specify a benchmark approach that relies on imputation and adjustment of data to reduce attrition in our analytic sample. We test this approach by comparing benchmark results with those produced by the same empirical model but with a reduced analytic sample that does not include cases that rely on imputed or adjusted data. Coefficients and p -values for the treatment indicator are presented in the tables below under Sensitivity Study 2. Again, as can be seen in Tables E.1 and E.2, the results produced with both analytic samples do not change inferential findings. The estimated treatment effects for both the benchmark and alternative reduced sample are not significant. Consequently, we infer that findings are not sensitive to the decision to impute or otherwise adjust missing data.

Sensitivity Study 3: Unreliable Data

In our benchmark analytic approach, we treat cases with what is deemed to be unreliable data as unit missing (see data cleaning section in Appendix F) and exclude them from the analytic sample. We test whether this analytic decision has an effect on substantive findings by comparing benchmark results with those produced by the same procedures, but with an analytic sample that includes the cases with unreliable data. Estimated treatment effects (coefficients and p -values) for this analytic sample are presented in Tables E.1 and E.2 tables under Sensitivity

Study 3. As can be seen, the results are inferentially similar. Estimated treatment effects are statistically insignificant for both models and both outcomes.

Sensitivity Study 4: Outliers

Our benchmark approach is to include all cases with observations that are identified as outliers (see data cleaning section of Appendix F). We test whether this analytic decision has an effect on inferential findings by comparing benchmark results with those produced by the same procedures but excluding the specific values that are identified as outliers (we convert all outliers to missing and then impute as we would other missing data). Coefficients and *p*-values for the treatment indicator are presented in the tables below under Sensitivity Study 4. As can be seen, the inclusion of outliers does not change inferential findings. For both outcomes, the coefficients for the treatment indicators are not significant in either the benchmark or sensitivity data.

Sensitivity Study 5: Condensed Programming

In the first cohort/summer, the BART intervention (and Healthy Living comparison intervention) was conducted over eight weeks, with one intervention session being conducted each week. In subsequent cohort/summers, the duration of the intervention was condensed from eight to six weeks (and in a few instances, 5 weeks) because the NYW program itself did not practicably permit an eight week program. For the condensed programming durations, the number of sessions was not reduced. Participants in both conditions still received each of the eight sessions in the same order; however, they received them over a shorter duration of time – six or five weeks. This fifth sensitivity study is included to determine whether or not the participants who received the condensed programming are differently impacted than those who received the intended eight-week-long intervention. We are interested in comparing the relative

effectiveness of the program for two distinct groups (i.e., condensed versus the full-term samples) rather than the relative effectiveness of the program for the full sample (i.e., benchmark that includes both condensed and full-term programming) compared to a subgroup of that sample as we do in the first four sensitivity studies. We do this by adding two variables to the benchmark analytic model: a variable for condensed programming (coded as 1 if a participant received the program in less than eight weeks and 0 otherwise) and an interaction term that is the product of the treatment indicator and the condensed programming indicator. Coefficients and p-values for the interaction term represent the differential effect of program duration for those in the treatment condition and are presented in the tables below under Sensitivity Study 5. As can be seen in Tables E.1 and E.2, the interaction term is not significant for either outcome (inconsistency of condom use and frequency of sex). Consequently, we infer that there is no significant differential effect of treatment for youth who participated in a full eight week program as compared to those who participated in a condensed course.

Sensitivity Study 6: Late Responders

Data collection windows were broad to minimize attrition from the analytic sample. To examine whether this influences our results – and, in particular, whether or not study participants who responded later report different outcomes from those who responded earlier – we conducted an analysis that compares impact estimates for treatment group participants who completed the questionnaire close to the open of their data collection to those of late responders. Late responders are defined as those participants who complete their six-month questionnaire more than two months after the initiation of the six-month data collection window. Late response time may influence treatment impacts for one of two reasons: a) Late response time may be indicative of a group characteristic, such as a low level of engagement with the program, that makes group

members less likely to be affected by and respond to the program's message, or b) Treatment impacts may be predicated upon time from exposure, meaning that late responders may show differential effects of treatment because of the length of time that had elapsed between the program ending and their response to the survey.

Again, since we are interested in comparing the relative effectiveness of the program for two distinct groups (i.e., early vs. late responders) rather than comparing the relative effectiveness of the program for the full sample (i.e., benchmark that includes both early and late responders) compared to a subgroup of that sample, the comparison of interest is not with the benchmark estimates but between each of the two (exclusive) groups. That is, we are assessing whether those who responded early and were exposed to the treatment exhibit significantly different results than those who responded late and were exposed to treatment. We do this by adding two variables to the benchmark analytic model: a variable for late responder (coded as 1 and 0 otherwise) and an interaction term that is the product of the treatment indicator and the late responder indicator. Coefficients and *p*-values for the interaction term represent the differential effect reported by late and early responders in the treatment condition. Results presented in the tables below under Sensitivity Study 6. Parameter estimates for the interaction terms are statistically insignificant for both outcomes of interests. Consequently, we infer that there is no significant differential effect of treatment for youth who respond late as compared to those who respond early.

Table E.1. Sensitivity of impact analyses using data from the 6-month follow-up questionnaire to address the primary research question

BART compared to control	Benchmark <i>b</i>	Benchmark <i>p</i>	Study 1 <i>b</i>	Study 1 <i>p</i>	Study 2 <i>b</i>	Study 2 <i>p</i>	Study 3 <i>b</i>	Study 3 <i>p</i>	Study 4 <i>b</i>	Study 4 <i>p</i>	Study 5 <i>b</i>	Study 5 <i>p</i>	Study 6 <i>b</i>	Study 6 <i>p</i>
Inconsistency of condom use	0.02	0.202	0.02	0.304	0.01	0.586	0.02	0.319	0.02	0.17	0.06	0.179	0.06	0.353

Source: Follow-up surveys administered 6 to 12 months after the program.

Notes: *b* refers to the regression adjusted mean difference in the outcome between BART and Healthy Living. *p* refers to the p-value of the difference; results are considered significant if $p < .05$. See Table III.3 for a more detailed description the outcome measure and section III for a description of the impact estimation methods.

Table E.2. Sensitivity of impact analyses using data from the 6-month follow-up questionnaire to address the secondary research question

BART compared to control	Benchmark <i>b</i>	Benchmark <i>p</i>	Study 1 <i>b</i>	Study 1 <i>p</i>	Study 2 <i>b</i>	Study 2 <i>p</i>	Study 3 <i>b</i>	Study 3 <i>p</i>	Study 4 <i>b</i>	Study 4 <i>p</i>	Study 5 <i>b</i>	Study 5 <i>p</i>	Study 6 <i>b</i>	Study 6 <i>p</i>
Frequency of sex	-0.17	0.653	-0.38	0.397	0.07	0.863	-0.07	0.837	-0.05	0.451	0.57	0.502	1.54	0.192

Source: Follow-up surveys administered 6 to 12 months after the program.

Notes: *b* refers to the regression adjusted mean difference in the outcome between BART and Healthy Living. *p* refers to the p-value of the difference; results are considered significant if $p < .05$. See Table III.3 for a more detailed description the outcome measure and section III for a description of the impact estimation methods.

Appendix F: Implementation evaluation methods

Table D.1. Methods used to address implementation research questions

Implementation element	Methods used to address each implementation element
<p>Adherence: How often were sessions offered? How many were offered?</p> <p>(Each total sessions offered statistic is reported overall and by cohort.)</p>	<p><i>Total number of BART sessions offered is a sum of the sessions captured by the Attendance Sheets.</i></p> <p><i>Average weekly frequency of BART sessions (by cohort) is calculated as the sum of the total number of sessions offered each week divided by the total number of active classes (per cohort). Statistics are reported for each of the possible eight sessions by cohort/year. Both numerator and denominator are captured by the Attendance Sheet.</i></p>
<p>Adherence: What and how much was received?</p> <p>(Each statistic will be reported overall and by cohort.)</p>	<p><i>Percentage of participants who attended each BART session (1-8) is calculated as the total number of participants who attended each session divided by the total number of participants assigned to the condition, as captured by the Attendance Sheet.</i></p> <p><i>Average number of BART sessions attended per participant is calculated as the sum of the total number of sessions attended by each participant divided by the total number of participants assigned to the BART condition. (Note: a participant may attend a maximum of eight sessions.)</i></p> <p><i>Percentage of treatment sample that attended all BART sessions is calculated as the total number of participants who attended all BART session divided by the total number of participants assigned to the BART condition.</i></p> <p><i>Percentage of treatment sample that did not attend any BART sessions is calculated as the total number of participants who failed to attend any BART session divided by the total number of participants assigned to the BART condition.</i></p>
<p>Adherence: What content was delivered to youth?</p> <p>(Each statistic is reported separately by reviewer type: health educator self-reports and observer (fidelity monitor) reports; we will also report ‘any’, which takes in account both reviewer types – the observer report if there is one and otherwise as reported by facilitator. Each statistic is reported overall and by cohort.)</p>	<p><i>Average number of BART intervention activities completed for each session type (session 1-8) is calculated as the sum of the total number of activities completed (for sessions one - eight separately) divided by the total number of sessions for which we have health educator self-reports/fidelity monitor observations, reported for each session type. (Note: There are 6 activities in session 1, 7 in session 2, 5 in session 3, 4 in session 4, 5 in session 5, 2 in session 6, 4 in session 7, and 4 in session 8. An activity is only considered complete if the health educator/fidelity monitor observer has marked “yes, completely” next to the activity on the BART Implementation Fidelity Tools.)</i></p> <p><i>Percentage of each type of BART session (1-8) in which 75% of intervention activities are completed is calculated as the total number of each type of session in which 75% of activities are completed, divided by the total number of each type of session for which we have health educator self-reports/fidelity monitor observations. (Note: we consider 75% of activities to be the following: 4 activities for session 1, 5 for session 2, 3 for session 3, 3 for session 4, 3 for session 5, 1 for session 6, 3 for session 7, and 3 for session 8.)</i></p> <p><i>Percentage of each type of BART session (1-8) in which 100% of intervention activities are completed is calculated as the total number of each type of session in which all activities are completed, divided by the total number of each type of session for which we have health educator self-reports/fidelity monitor observations.</i></p>

Implementation element	Methods used to address each implementation element
<p>Adherence: Who delivered material to youth?</p>	<p><i>Average number of BART intervention activities completed for each session type (session 1-8) is calculated as the sum of the total number of activities completed (for sessions one - eight separately) divided by the total number of sessions for which we have health educator self-reports/fidelity monitor observations, reported for each session type. (Note: There are 6 activities in session 1, 7 in session 2, 5 in session 3, 4 in session 4, 5 in session 5, 2 in session 6, 4 in session 7, and 4 in session 8. An activity is only considered complete if the health educator/fidelity monitor observer has marked “yes, completely” next to the activity on the BART Implementation Fidelity Tools.)</i></p> <p><i>Percentage of each type of BART session (1-8) in which 75% of intervention activities are completed is calculated as the total number of each type of session in which 75% of activities are completed, divided by the total number of each type of session for which we have health educator self-reports/fidelity monitor observations. (Note: we consider 75% of activities to be the following: 4 activities for session 1, 5 for session 2, 3 for session 3, 3 for session 4, 3 for session 5, 1 for session 6, 3 for session 7, and 3 for session 8.)</i></p> <p><i>Percentage of each type of BART session (1-8) in which 100% of intervention activities are completed is calculated as the total number of each type of session in which all activities are completed, divided by the total number of each type of session for which we have health educator self-reports/fidelity monitor observations.</i></p>
<p>Quality: Quality of staff-participant interactions</p> <p>(Each statistic is reported overall and by cohort, for both treatment and counterfactual groups.)</p>	<p><i>Percentage of observed sessions where the fidelity monitor scored the delivery of session information to participants as “good” (=4) or “very good” (=5) is calculated as the total number of sessions for which the average score for questions 1-3 from the Program Observation Form for TPP Grantees = 4 or 5, divided by the total number of observed sessions. (Delivery of session information is a scale variable that is constructed as the average score of item responses to questions 1-3 in the Program Observation Form for TPP Grantees. Response options for questions 1-3 range from 1-5, with 1 being the worst rating and 5 being the best rating. For each rated session, the scale score could range from 1=very poor delivery of information to 5=very good delivery of information).</i></p> <p><i>Percentage of observed sessions where the fidelity monitor scored the extent of participants’ understanding of session material as “moderate” (=4) or “good” (=5) is calculated as the total number of sessions for which the score for question 4 in the Program Observation Form for TPP Grantees = 4 or 5, divided by the total number of observed sessions. (Extent of participants’ understanding is operationalized as the response to question 4 in the Program Observation Form for TPP Grantees; response items range from 1=little understanding to 5=good understanding).</i></p> <p><i>Percentage of observed sessions where the fidelity monitor scored the level of group participation in session discussions and activities as “moderate” (=4) or “active” (=5) is calculated as the total number of sessions for which the score for question 5 in the Program Observation Form for TPP Grantees = 4 or 5, divided by the total number of observed sessions. (Level of group participation is operationalized as the response to question 5 from Program Observation Form for TPP Grantees; response items range from 1=little participation to 5=active participation).</i></p> <p><i>Percentage of observed sessions where the fidelity monitor scored the overall quality of the program session as “very good” (=4) or “excellent” (=5) is calculated as the total number of sessions for which the score for question 7 in the Program Observation Form for TPP Grantees = 4 or 5, divided by the total number of observed sessions. (Overall quality of program session is operationalized as the response to question 7 from Program Observation Form for TPP Grantees; response items range from 1=poor to 5=excellent).</i></p>

Implementation element	Methods used to address each implementation element
Quality: Quality of youth engagement with program	<i>A benchmark of the quality of youth engagement is calculated as the percentage of sessions in which the independent evaluator scored youth engagement as “moderately engaged” (4) or higher.</i>
Counterfactual: Experiences of counterfactual condition (Each total sessions offered statistic above is reported overall and by cohort.)	<i>Total number of Healthy Living sessions (1-8) offered is a sum of the session 1s offered captured by the Attendance Sheet. Average weekly frequency of Healthy Living sessions (by cohort) is calculated as the sum of the total number of sessions offered each week during each cohort divided by the total number of active classes (per cohort). Statistics is reported for each of the possible eight sessions by cohort/year. Both numerator and denominator are captured by the Attendance Sheet.</i>
Counterfactual: What and how much was received? (Each statistic is reported overall and by cohort.)	<i>Percentage of participants who attended each Healthy Living session (1-8) is calculated as the total number of participants who attended each session divided by the total number of participants assigned to the condition, captured by Attendance Sheet. Average number of Healthy Living sessions attended per participant is calculated as the sum of the total number of sessions attended by each participant divided by the total number of participants assigned to the condition. (Note: a participant may attend a maximum of eight sessions.) Percentage of counterfactual sample that attended all Healthy Living sessions is calculated as the total number of participants who attended all 8 session divided by the total number of participants assigned to the condition. Percentage of counterfactual sample that did not attend any Healthy Living sessions is calculated as the total number of participants who failed to attend any session divided by the total number of participants assigned to the condition.</i>

Appendix G: Implementation evaluation results

Adherence

How many and how often were sessions offered?

Table G.1. Total number of BART sessions offered, by cohort and overall

Cohort	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
2012	16	16	16	16	16	16	16	16
2013	14	14	14	14	14	14	12	12
2014	13	13	13	13	13	13	13	13
All cohorts	43	43	43	43	43	43	41	41

Note: sessions that were not offered were cancelled/not held due to low/no attendance or because classrooms were unexpectedly unavailable at program sites. The total number of sessions that should have been offered across the 43 classes held was 344

Table G.2. Number of BART classes for which programming was offered over a period of five, six, seven, and eight weeks

Cohort	5 weeks	6 weeks	7 weeks	8 weeks
2012	0	0	7	9
2013	6	8	0	0
2014	0	13	0	0
All cohorts	6	21	7	9

Table G.3. Total number of BART classes and average number of weeks of programming and sessions offered per week

Cohort	Number of classes	Average weeks of programming	Average number of sessions offered per week
2012	16	7.6	1.1
2013	14	5.6	1.4
2014	13	6	1.3
All cohorts	43	6.4	1.2

What and how much was received?

Table G.4. Percentage of participants who attended each BART session, by cohort and overall

Cohort	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
2012 (n=164)	91.5	86.6	82.3	80.5	85.4	81.7	72.6	69.5
2013 (n=148)	79.7	85.1	79.1	83.8	79.1	77.0	75.0	70.9
2014 (n=115)	81.7	80.9	78.3	78.3	69.6	72.2	70.4	66.1
All cohorts (n=427)	84.8	84.5	80.1	81.0	78.9	77.5	72.8	69.1

Table G.5. Average number of BART sessions attended per participant, by cohort and overall

Cohort	Number of participants	Average number of sessions
2012	164	6.5
2013	148	6.3
2014	115	6.0
All cohorts	427	6.3

Table G.6. Percentage of participants who attended no and all BART sessions, by cohort and overall

Cohort	Number of participants	Percent attended no sessions	Percent attended all sessions
2012	164	0.6	40.2
2013	148	2.0	40.5
2014	115	7.0	39.1
All cohorts	427	2.8	40.0

What content was delivered to youth?

Table G.7. Average number of intervention activities completed for each BART session, by cohort and overall

Cohort	Review type	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
2012	HE	4.4	6.1	3.8	3.6	3.8	2.0	3.6	3.3
	Obs	4.5	5.3	5.0	4.0	4.5	2.0	4.0	4.0
	Any	4.4	5.9	4.0	3.8	3.8	2.0	3.7	3.3
2013	HE	5.0	6.8	4.7	3.9	4.4	1.8	2.9	3.5
	Obs	4.4	4.0	3.3	4.0	4.0	2.0	3.3	4.0
	Any	5.1	6.6	4.5	3.9	4.3	1.8	2.8	3.5
2014	HE	4.1	6.5	4.8	4.0	4.7	1.9	3.9	4.0
	Obs	4.7	5.7	4.3	4.0	4.0	1.8	3.4	4.0
	Any	4.1	6.2	4.6	4.0	4.5	1.8	3.8	4.0
All cohorts	HE	4.5	6.4	4.3	3.8	4.2	1.9	3.5	3.5
	Obs	4.5	5.3	4.3	4.0	4.2	1.9	3.5	4.0
	Any	4.5	6.2	4.4	3.9	4.2	1.9	3.5	3.6
<i>Intended activities</i>		6	7	5	4	5	2	4	4
<i>Average percent completed</i>		75.0%	88.6%	88.0%	97.5%	84.0%	95.0%	87.5%	90.0%

Note: For reviewer type: HE = health educator self-reports; Obs = fidelity monitor observer reports; and Any = report taking into account both - observer report taken if there is one, otherwise as reported by facilitator. *Average percent of activities completed* is calculated as the quotient of the “any” reviewer average number of sessions for all cohorts divided by the intended number of activities.

Note: There are 6 activities in session 1, 7 in session 2, 5 in session 3, 4 in session 4, 5 in session 5, 2 in session 6, 4 in session 7, and 4 in session 8. An activity is only considered complete if the health educator/fidelity monitor observer has marked “yes, completely” next to the activity on the BART Implementation Fidelity Tools.

Limitations note: 1) health educator self-reports may not be a reliable measure of the content that was actually delivered to participants; additionally, we do not have complete self-report data for all BART intervention sessions delivered; we have self-report data for: 95% (41/43) of session 1s, 91% (39/43) of session 2s, 95% (41/43) of session 3s, 98% (42/43) of session 4s, 98% (42/43) of session 5s, 95% (41/43) of session 6s, 91% (39/43) of session 7s, and 70% (30/43) of session 8s; 2) observer data are very incomplete and may thus fail to offer a representative picture of the content actually delivered to youth; we have limited observation data for all BART session types - we have observation data for: 23% (10/43) of session 1s, 19% (8/43) of session 2s, 33% (14/43) of session 3s, 21% (9/43) of session 4s, 23% (10/43) of session 5s, 21% (9/43) of session 6s, 28% (12/43) of session 7s, and 12% (5/43) of session 8s.

Table G.8. Percentage of BART sessions in which 75% of intervention activities were completed, by cohort and overall

Cohort	Review type	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
2012	HE	75.0	87.5	87.5	87.5	81.3	100.0	93.8	80.0
	Obs	50.0	75.0	100.0	100.0	100.0	100.0	100.0	100.0
	Any	68.8	87.5	87.5	93.8	81.3	100.0	93.8	80.0
2013	HE	84.6	100.0	100.0	100.0	100.0	91.7	70.0	87.5
	Obs	80.0	0.0	66.7	100.0	100.0	100.0	75.0	100.0
	Any	92.3	91.7	92.3	100.0	100.0	91.7	63.6	87.5
2014	HE	58.3	90.9	100.0	100.0	100.0	100.0	100.0	100.0
	Obs	66.7	66.7	100.0	100.0	66.7	100.0	80.0	100.0
	Any	58.3	81.8	100.0	100.0	92.3	100.0	92.3	100.0
All cohorts	HE	73.2	92.3	95.1	95.2	92.9	97.6	89.7	86.7
	Obs	70.0	62.5	92.9	100.0	90.0	100.0	83.3	100.0
	Any	73.2	87.2	92.9	97.6	90.5	97.6	85.0	87.9

Note: For reviewer type: HE = health educator self-reports; Obs = fidelity monitor observer reports; and Any = report taking into account both - observer report taken if there is one, otherwise as reported by facilitator.

Note: we consider 75% of activities to be the following: 4 activities for session 1, 5 for session 2, 3 for session 3, 3 for session 4, 3 for session 5, 1 for session 6, 3 for session 7, and 3 for session 8. An activity is only considered complete if the health educator/fidelity monitor observer has marked "yes, completely" next to the activity on the BART Implementation Fidelity Tools.

Limitations note: 1) health educator self-reports may not be a reliable measure of the content that was actually delivered to participants; additionally, we do not have complete self-report data for all BART intervention sessions delivered; we have self-report data for: 95% (41/43) of session 1s, 91% (39/43) of session 2s, 95% (41/43) of session 3s, 98% (42/43) of session 4s, 98% (42/43) of session 5s, 95% (41/43) of session 6es, 91% (39/43) of session 7s, and 70% (30/43) of session 8s; 2) observer data are very incomplete and may thus fail to offer a representative picture of the content actually delivered to youth; we have limited observation data for all BART session types - we have observation data for: 23% (10/43) of session 1s, 19% (8/43) of session 2s, 33% (14/43) of session 3s, 21% (9/43) of session 4s, 23% (10/43) of session 5s, 21% (9/43) of session 6es, 28% (12/43) of session 7s, and 12% (5/43) of session 8s.

Table G.9. Percentage of BART sessions in which 100% of intervention activities were completed, by cohort and overall

Cohort	Review type	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
2012	HE	37.5	56.3	25.0	75.0	37.5	100.0	62.5	60.0
	Obs	50.0	25.0	100.0	100.0	75.0	100.0	100.0	100.0
	Any	43.8	50.0	43.8	87.5	43.8	100.0	75.0	60.0
2013	HE	46.2	83.3	69.2	92.3	53.8	83.3	40.0	75.0
	Obs	20.0	0.0	33.3	100.0	33.3	100.0	50.0	100.0
	Any	46.2	83.3	76.9	92.3	46.2	83.3	36.4	75.0
2014	HE	33.3	72.7	83.3	100.0	69.2	92.3	92.3	100.0
	Obs	66.7	66.7	57.1	100.0	66.7	80.0	60.0	100.0
	Any	33.3	63.6	76.9	100.0	69.2	84.6	84.6	100.0
All cohorts	HE	39.0	69.2	56.1	88.1	52.4	92.7	66.7	73.3
	Obs	40.0	37.5	64.3	100.0	60.0	88.9	66.7	100.0
	Any	41.5	64.1	64.3	92.9	52.4	90.2	67.5	75.8

Note: For reviewer type: HE = health educator self-reports; Obs = fidelity monitor observer reports; and Any = report taking into account both - observer report taken if there is one, otherwise as reported by facilitator.

Note: 100% of intervention activities is the following: 6 activities in session 1, 7 in session 2, 5 in session 3, 4 in session 4, 5 in session 5, 2 in session 6, 4 in session 7, and 4 in session 8. An activity is only considered complete if the health educator/fidelity monitor observer has marked "yes, completely" next to the activity on the BART Implementation Fidelity Tools.

Limitations note: 1) health educator self-reports may not be a reliable measure of the content that was actually delivered to participants; additionally, we do not have complete self-report data for all BART intervention sessions delivered; we have self-report data for: 95% (41/43) of session 1s, 91% (39/43) of session 2s, 95% (41/43) of session 3s, 98% (42/43) of session 4s, 98% (42/43) of session 5s, 95% (41/43) of session 6es, 91% (39/43) of session 7s, and 70% (30/43) of session 8s; 2) observer data are very incomplete and may thus fail to offer a representative picture of the content actually delivered to youth; we have limited observation data for all BART session types - we have observation data for: 23% (10/43) of session 1s, 19% (8/43) of session 2s, 33% (14/43) of session 3s, 21% (9/43) of session 4s, 23% (10/43) of session 5s, 21% (9/43) of session 6es, 28% (12/43) of session 7s, and 12% (5/43) of session 8s.

Who delivered material to youth?

LPHI hired health educators to implement the intervention. Teams consisting of two health educators (one male and one female) were responsible for leading the BART and Healthy Living interventions within the group setting according to their respective curricula, record attendance, complete fidelity monitoring instruments, and be available for any scheduled make-up sessions, should they be necessary. All health educators had to meet the following qualifications:

- High school graduate, with some college experience in Public Health, and/or Health Education, or Education; advanced degree strongly preferred; teaching experience is a plus
- Strong interpersonal skills and strong organizational skills.
- Excellent communication skills
- Professional attitude and manner that reflects the high standards of program & sensitive topics
- Some experience interacting with adolescents and strong commitment to youth education

- Ability to manage many interrelated tasks at once.
- Computer skills with Microsoft Office Programs
- Genuine sensitivity to the needs of all children and commitment to youth education

Table G.10. Total number of health educators who facilitated BART and Healthy Living sessions, by cohort and overall

Cohort	Number of health educators
2012	16
2013	14
2014	20
Total	41

Note: Rows do not sum to total because health educators could serve in more than one year.

Table G.11. Percentage of health educators trained in BART curriculum, Healthy Living curriculum, Fidelity Monitoring, and Evaluation Research Basics

Cohort	Bart Curriculum Training	Healthy Living Curriculum Training	Fidelity Monitoring Training	Evaluation Research Basics Training	Completed all four trainings
2012 (n=16)	16	16	13	13	81.3%
2013 (n=14)	14	14	14	14	100.0%
2014 (n=20)	20	20	17	17	85%
Total (n=41)	41	41	36	36	87.8%

Note: Rows do not sum to total because health educators could serve in more than one year.

Quality of Staff-Participant Interactions

Table G.12. Percentage of observed BART and Healthy Living sessions in which staff-participant interactions were rated good/moderate or better by fidelity monitor observers, by cohort and overall

Cohort	Delivery of session information (scored good or very good)	Extent of participants' understanding (scored moderate or good)	Extent of group members' participation (scored moderate or active)	Overall quality of the program session (scored good or excellent)
BART				
2012 (n=25)	60.0%	72.0%	72.0%	60.0%
2013 (n=20)	30.0%	60.0%	60.0%	60.0%
2014 (n=35)	85.7%	74.3%	62.9%	71.4%
Total (n=80)	63.8%	70.0%	65.0%	65.0%
Healthy Living				
2012 (n=30)	66.7%	53.3%	83.3%	50.0%
2013 (n=22)	68.2%	54.5%	54.5%	52.4%
2014 (n=23)	78.3%	69.6%	82.6%	78.3%
Total (n=75)	70.7%	58.7%	74.7%	59.5%

Note: Data that are used to assess quality of staff-participant interactions are not representative of all interactions; they are based on a limited convenience sample of observed sessions. In all 86 classes were held (43 BART and 43 Healthy Living), with an intended 8 sessions each; therefore data were gathered on approximately 23% of all sessions (155 observed/ 688 intended).

Note: For all quality of staff-participant interactions measures, response options range from 1-5, with 1 being the worst rating and 5 being the best rating. **Delivery of session information** is a scale variable constructed from questions 1-3 from the *Program Observation Form for TPP Grantees* and calculated as the percentage of observed sessions where the average score is "good" (=4) or "very good" (=5). **Extent of participants' understanding** is calculated as the percentage of observed sessions where the fidelity monitor scored question 4 in the *Program Observation Form for TPP Grantees* as "moderate" (=4) or "good" (=5). **Level of group participation** is calculated as the percentage of observed sessions where the fidelity monitor scored question 5 from *Program Observation Form for TPP Grantees* as "moderate" (=4) or "active" (=5). **Overall quality of program session** is calculated as the percentage of observed sessions where the fidelity monitor scored question 7 from the *Program Observation Form for TPP Grantees* as "very good" (=4) or "excellent" (=5).

Counterfactual

How many and how often were sessions offered?

Table G.13. Total number of Healthy Living sessions offered, by cohort and overall

Cohort	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
2012	16	16	16	16	16	16	16	15
2013	14	13	14	14	14	14	12	12
2014	13	13	13	13	13	13	12	13
All cohorts	43	42	43	43	43	43	40	40

Note: sessions that were not offered were cancelled/not held due to low/no attendance or because classrooms were unexpectedly unavailable at program sites. The total number of sessions that should have been offered across the 43 classes held was 344

Table G.14. Number of Healthy Living classes for which programming was offered over a period of five, six, seven, and eight weeks

Cohort	5 weeks	6 weeks	7 weeks	8 weeks
2012	0	2	4	10
2013	6	8	0	0
2014	0	13	0	0
All cohorts	6	23	4	10

Table G.15. Total number of Healthy Living classes and average number of weeks of programming and sessions offered per week

Cohort	Number of classes	Average weeks of programming	Average number of sessions offered per week
2012	16	7.6	1.1
2013	14	5.6	1.4
2014	13	6	1.3
All cohorts	43	6.4	1.2

What and how much was received?

Table G.16. Percentage of participants who attended each Healthy Living session, by cohort and overall

Cohort	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
2012 (n=170)	87.1	88.2	85.3	74.1	82.4	75.9	68.8	67.1
2013 (n=139)	79.9	80.6	86.3	84.9	77.7	78.4	78.4	73.4
2014 (n=114)	80.7	79.8	74.6	75.4	71.1	64.0	62.3	65.8
All cohorts (n=423)	83.0	83.5	82.7	78.0	77.8	73.5	70.2	68.8

Table G.17. Average number of Healthy Living sessions attended per participant, by cohort and overall

Cohort	Number of participants	Average number of sessions
2012	170	6.3
2013	139	6.4
2014	114	5.7
All cohorts	423	6.2

Table G.18. Percentage of participants who attended no and all Healthy Living sessions, by cohort and overall

Cohort	Number of participants	Percent attended no sessions	Percent attended all sessions
2012	170	2.4	35.3
2013	139	4.3	38.8
2014	114	5.3	34.2
All cohorts	423	3.8	36.2

What content was delivered to youth?

Table G.19. Average number of intervention activities completed for Healthy Living session 1, by cohort and overall

Cohort	Review type	Number of observations	Average number of activities completed
2012	HE	16	4.6
	Obs	4	5.8
	Any	16	4.6
2013	HE	13	4.6
	Obs	2	5.0
	Any	13	4.6
2014	HE	13	3.9
	Obs	7	5.4
	Any	13	4.2
All cohorts	HE	42	4.4
	Obs	13	5.5
	Any	42	4.5

Note: Healthy Living session 1 is identical to BART session 1. There are six activities in session 1; an activity is only considered complete if the health educator/fidelity monitor has marked “yes, completely” next to the activity on the Healthy Living Implementation Fidelity Tools.

Note: For reviewer type: HE = health educator self-reports; Obs = fidelity monitor observer reports; and Any = report taking into account both - observer report taken if there is one, otherwise as reported by facilitator.

Limitations note: 1) health educator self-reports may not be a reliable measure of the content that was actually delivered to participants; additionally, we do not have complete self-report data for all Healthy Living counterfactual intervention sessions delivered; we have self-report data for: 98% (42/43) of session 1s, 93% (40/43) of session 2s, 98% (42/43) of session 3s, 93% (40/43) of session 4s, 98% (42/43) of session 5s, 98% (42/43) of session 6es, 93% (40/43) of session 7s, and 72% (31/43) of session 8s; 2) observer data are very incomplete and may thus fail to offer a representative picture of the content actually delivered to youth; we have limited observation data for all Healthy Living counterfactual session types; we have observation data for: 30% (13/43) of session 1s, 30% (13/43) of session 2s, 19% (8/43) of session 3s, 19% (8/43) of session 4s, 12% (5/43) of session 5s, 33% (14/43) of session 6es, 16% (7/43) of session 7s, and 21% (9/43) of session 8s.

Table G.20. Percentage of Healthy Living session 1s in which 75% and 100% of intervention activities were completed, by cohort and overall

Cohort	Review type	Number of observations	Percent 75% complete	Percent 100% complete
2012	HE	16	75.0	43.8
	Obs	4	100.0	75.0
	Any	16	75.0	50.0
2013	HE	13	84.6	30.8
	Obs	2	100.0	50.0
	Any	13	84.6	30.8
2014	HE	13	53.8	30.8
	Obs	7	85.7	85.7
	Any	13	53.8	53.8
All cohorts	HE	42	71.4	35.7
	Obs	13	92.3	76.9
	Any	42	71.4	45.2

Note: Healthy Living session 1 is identical to BART session 1. There are six activities in session 1; an activity is only considered complete if the health educator/fidelity monitor has marked "yes, completely" next to the activity on the Healthy Living Implementation Fidelity Tools.

Note: For reviewer type: HE = health educator self-reports; Obs = fidelity monitor observer reports; and Any = report taking into account both - observer report taken if there is one, otherwise as reported by facilitator.

Limitations note: 1) health educator self-reports may not be a reliable measure of the content that was actually delivered to participants; additionally, we do not have complete self-report data for all Healthy Living counterfactual intervention sessions delivered; we have self-report data for: 98% (42/43) of session 1s, 93% (40/43) of session 2s, 98% (42/43) of session 3s, 93% (40/43) of session 4s, 98% (42/43) of session 5s, 98% (42/43) of session 6es, 93% (40/43) of session 7s, and 72% (31/43) of session 8s; 2) observer data are very incomplete and may thus fail to offer a representative picture of the content actually delivered to youth; we have limited observation data for all Healthy Living counterfactual session types; we have observation data for: 30% (13/43) of session 1s, 30% (13/43) of session 2s, 19% (8/43) of session 3s, 19% (8/43) of session 4s, 12% (5/43) of session 5s, 33% (14/43) of session 6es, 16% (7/43) of session 7s, and 21% (9/43) of session 8s.

Table G.21. Percentage of Healthy Living sessions two through eight in which the health educator engaged in any (one or more) of the six core components of BART

Cohort	Review type	Session 2 percent (obs)	Session 3 percent (obs)	Session 4 percent (obs)	Session 5 percent (obs)	Session 6 percent (obs)	Session 7 percent (obs)	Session 8 percent (obs)
2012	HE	50 (16)	6.3 (16)	0 (16)	0 (16)	0 (16)	0 (16)	0 (15)
	Obs	50 (4)	0 (4)	0 (2)	0 (3)	0 (4)	0 (4)	0 (5)
	Any	50 (16)	6.3 (16)	0 (16)	0 (16)	0 (16)	0 (16)	0 (15)
2013	HE	8.3 (12)	0 (13)	9.1 (11)	0 (13)	0 (13)	0 (12)	0 (9)
	Obs	0 (6)	0 (3)	0 (5)	0 (1)	0 (4)	0 (1)	0 (1)
	Any	8.3 (12)	0 (14)	8.3 (12)	0 (13)	0 (14)	0 (12)	0 (10)
2014	HE	0 (12)	0 (13)	0 (13)	7.7 (13)	0 (13)	0 (12)	0 (7)
	Obs	0 (3)	0 (1)	0 (1)	0 (1)	0 (6)	0 (2)	0 (3)
	Any	0 (12)	0 (13)	0 (13)	7.7 (13)	0 (13)	0 (12)	0 (10)
All cohorts	HE	22.5 (40)	2.4 (42)	2.5 (40)	2.4 (42)	0 (42)	0 (40)	0 (31)
	Obs	15.4 (13)	0 (8)	0 (8)	0 (5)	0 (14)	0 (7)	0 (9)
	Any	22.5 (40)	2.3 (43)	2.4 (41)	2.4 (42)	0 (43)	0 (40)	0 (35)
Sessions in which knowledge discussed	Any	17.5 (40)	0 (43)	2.4 (41)	2.4 (42)	0 (43)	0 (40)	0 (35)
Sessions in which other core content discussed	Any	5 (40)	2.3 (43)	0 (41)	0 (42)	0 (43)	0 (40)	0 (35)

Note: There are six core components of BART – HIV/AIDS knowledge, ways to handle social and sexual pressures, ways to communicate assertively about sex, refusal skills related to sex, negotiation skills related to sex, and condom use skills. Any BART core component is considered “engaged in” for counterfactual sessions two through eight if the health educator/fidelity monitor marked “yes” on the Healthy Living Implementation Fidelity Tool next to any of the six core components listed for that session. Since HIV/AIDS knowledge was explicitly provided in session 1 and health educators were not instructed to avoid knowledge-related discussions in other sessions, we also break down findings according to sessions in which knowledge was discussed and sessions in which other core content was discussed – percentages are calculated as the number of sessions “any” reviewer indicated core content was engaged in, divided by all sessions reviewed.

Note: For reviewer type: HE = health educator self-reports; Obs = fidelity monitor observer reports; and Any = report taking into account both - observer report taken if there is one, otherwise as reported by facilitator. Obs in session columns = number of observations and % = percentage of observations in which the health educator engaged in any (one or more) of the six core components of BART.

Limitations note for Table G.20 also applies to this table.

Context

Table G.22. List of other known teen pregnancy prevention programming being implemented in Orleans Parish during the program period

Program name	Program lead agency	Funder/grantee type	City/State
Making Proud Choices (MPC!) (also known as: Believe in Youth! or BY!-NOLA!)	Institute of Women and Ethnic Studies	OAH - TPP-Tier 1	New Orleans, LA
Teen Outreach Program (TOPs Clubs)	Louisiana DHH Office of Public Health	OAH - TPP-Tier 1	New Orleans, LA
e-SIHLE	Tulane University	OAH - TPP-Tier 2	New Orleans, LA
Safer Sex (or Staying Mature and Responsible Towards Sex - SMARTS)	Louisiana Public Health Institute	OAH - TPP-Tier 1	New Orleans, LA
Project AIM (Adult Identity Mentoring) - adaptation	Louisiana Office of Public Health	OAH - PREP	New Orleans, LA
Focus on Your Future	University of Kentucky College of Public Health	NIMH – R01 Study	New Orleans, LA
<i>Be Proud! Be Responsible! (BPBR)</i>	<i>Central Louisiana Area Health Education Center Foundation</i>	<i>OAH - TPP-Tier 1</i>	<i>Alexandria, LA</i>
<i>SIHLE</i>	<i>Louisiana Office of Public Health</i>	<i>OAH – PREP</i>	<i>Louisiana Regions 2-9</i>

Note: BPBR and SIHLE are included at the bottom in italics in this table because they were implemented during the program period, but outside of Orleans Parish.

Table G.23. Percentage of participants self-reporting past-year exposure to reproductive health education at each data collection point, overall and by treatment and comparison group

Cohort	All observations	All percent	BART observations	BART percent	Healthy Living observations	Healthy Living percent
Baseline	781	55.6	392	53.1	389	58.1
Post-program	735	60.7	370	61.6	365	59.7
6 months post-program	691	67.0	349	73.4	342	60.5

Note: The question asks “In the past year, please tell us if you have had any formal education classes in school or some other place, such as a community center, church, or health clinic, on any of the following: (Please choose ALL that apply).” The response options are: the female menstrual cycle (period); how pregnancy occurs; sexually transmitted infections (STIs); how to say “NO” to sex; methods of birth control – that is, how to stop a pregnancy from happening; how to prevent HIV/AIDS using safe sex practices; I have never had any formal educational classes on any of the above topics. Though post-program data are not used in this impact evaluation, they are included here as they potentially provide context to the 6-month post-program outcomes.

Table G.24. Percentage of participants self-reporting past-year experiences with one or more other teen pregnancy prevention programs at each data collection point, overall and by treatment and comparison group

Cohort	All observations	All percent	BART observations	BART percent	Healthy Living observations	Healthy Living percent
Baseline	747	14.9	371	16.4	376	13.3
Post-program	719	17.1	361	15.0	358	19.3
6 months post-program	708	9.7	356	9.8	352	9.7

Note: the question asks “In the past year, have you been a participant in any of the following youth programs? (Please choose ALL that apply).” The response options are: Becoming a Responsible Teen (BART); Healthy Living; 4 Real Health; Be Proud! Be Responsible!; MPC! – NOLA (Making Proud Choice – New Orleans, LA); Teen Outreach Program (also known as TOPs Clubs); Safer Sex; SMARTS (Staying Mature and Responsible Toward Sex); Sisters Informing, Healing, Living, and Empowering (SiHLE); Project AIM (Adult Identity Mentoring); Focus on Your Future!; Other(s) – (write in option); I have never been a participant in any of the youth programs listed above.

Note: Though HEP program names (BART, Healthy Living, and 4 Real Health) were included as response options for this question, participants who selected any of these three options were not counted as having an experience with an ‘other TPP’ program in the above table. However, it should be noted that although participants were diligently screened by study staff for prior participation in HEP before being enrolled in the study, at baseline, 90 youth (51 assigned to BART and 39 assigned to Healthy Living) self-reported on the questionnaire that they had participated in BART, Healthy Living, or 4 Real Health. Although this is concerning, we recognize that self-reports are often unreliable and, though the question asked about their participation in these programs in the past year, it’s possible that at least a portion of these youth answered affirmatively based on the fact that they were currently enrolled in 4Real Health (though they had not yet received their first program session at the time the baseline questionnaire was administered).