Evaluation of the Teen Outreach Program® in Rochester, New York

Final Impact Report for

City of Rochester
Department of Recreation and Youth Services

April, 2016

Prepared by

Hugh F. Crean, Ph.D.
Susan M. Seibold-Simpson, Ph. D.
Marc Jambon
Richard E. Kreipe, M.D.

Crean, H. F., Seibold-Simpson, S. M., Jambon, M., & Kreipe, R. E. (2015). *Evaluation of the Teen Outreach Program® in Rochester, New York: Finding from the replication of an evidence-based teen pregnancy prevention program*. Rochester, New York: University of Rochester School of Nursing.

EVALUATION OF THE TEEN OUTREACH PROGRAM®
IN ROCHESTER, NEW YORK:
FINDINGS FROM THE REPLICATION OF AN EVIDENCE-BASED
TEEN PREGNANCY PREVENTION PROGRAM

I. Introduction

Reducing adolescent pregnancy remains a priority for the United States. While rates have been decreasing steadily since the 1990s, important disparities remain, particularly for youth of color and youth living in poverty. Reducing adolescent pregnancy and child-bearing is important for several reasons—childbearing during adolescence negatively affects the parents, their children, and society. Compared with their peers who delay childbearing, adolescent girls who have babies are less likely to finish high school, more likely to rely on public assistance, more likely to be poor as adults, and their children are more likely to have poorer educational, behavioral, and health outcomes over the course of their lives than do children born to older parents (Coyne & D'Onofrio, 2012; Hoffman & Maynard, 2008). There is a clear need to provide effective, evidence-based programs designed to reduce rates of teen pregnancy and unsafe sexual behaviors. The opportunity to provide an intervention that also enhances positive youth development is an additional bonus that benefits adolescents and their communities as a whole.

Youth who live in Rochester, a city of approximately 200,000 people in upstate New York, have been plagued by poverty, violence, and poor educational achievement for decades. Only three major U.S. cities (Detroit, Cleveland, and Dayton, OH) have higher child poverty rates (Doherty, 2015). Significant racial and ethnic segregation in this urban center is reflected by an overwhelmingly African American and Hispanic population in the City of Rochester with a primarily white population in the surrounding county. Disparities between City youth and those in rest of the County are evident in a myriad of characteristics, including income, education, health, safety, and well-being. Perhaps not surprisingly given the above, the City of Rochester has consistently had the some of the highest rates of adolescent pregnancy within both the State of New York and the country and is often in the lower tier of children and adolescent health outcomes within the country. As of 2008, the birth rate for females age 15-19 in the City of Rochester was 80 per 1,000 (3 per 1,000 for females aged 10-14) (Metro Council for Teen Potential, 2012), well above the 41.5 per 1000 rate for that age group in the nation as a whole (Martin et al., 2010). A full 20% of the babies born in Rochester (681 of 3,409) in 2008 were born to teen mothers.

A. Introduction and study overview

In 2010, the U.S. Department of Health and Human Services' Office of Adolescent Health (OAH) awarded funds to the City of Rochester, along with approximately 100 other agencies and investigators, to replicate with fidelity and rigorously evaluate select evidence-based teen pregnancy prevention programs to both expand the research base and to update evidence in different settings and with a different generation of adolescents. These funds were used to create Rochester's Project THRIVE (Teens Helping Reinvent Identity, Values, and Empowerment). Rochester selected Wyman's Teen Outreach Program® (TOP®) from the more than 30 programs that met OAH review standards. TOP® is a 9-month positive youth development program that moves beyond basic sexuality education to provide structured small-group activities, caring adult

support and guidance, and community-service learning. At the time of study initiation, consistent with its positive youth development approach, TOP® had been shown effective in both increasing school success and preventing teen pregnancy in school settings (Allen & Philliber, 2001; Allen, Philliber, Herrling, & Kuperminc, 1997; Allen, Philliber, & Hoggson, 1990). Despite this, there are a number of concerns with this evidence: (1) the evidence is dated as the original trial is now close to 20 years old (Allen et al., 1997), (2) the sample from this trial was predominantly female (85% and the program was originally designed for adolescent females), and (3) the statistical analyses in the original trial ignored the clustering effect of students within schools (which were the unit of randomization). While several studies have examined TOP® in school settings, Rochester evaluated TOP® in community settings, specifically out-of-school programs hosted in school-based or free-standing urban recreation centers. TOP® is particularly relevant to community settings as it may keep youth engaged in constructive out-of-school activities, as opposed to more risky adolescent use of free time. Additionally, providing TOP® in community settings may enhance connectedness to the community (through CSL and other activities) and provide youth with access and opportunity to connect with caring and supportive adults in their community. Further, Rochester focused on the 11-14 year old age group; existing evidence is primarily with older, high-school aged youth using samples that were more racially and ethnically homogenous compared to the current sample.

B. Primary research question

The current evaluation tested the extent to which TOP®, when replicated with fidelity, produced impacts on sexual activity in the short-term (immediate post-intervention assessment period). The research questions were pre-specified and categorized as primary (to establish the effectiveness of the program) and secondary (additional questions about sexual intentions to provide evidence suggestive of program impact with a younger sample). The primary research question was:

- What is the impact of the TOP® relative to a Work Readiness (WR) curriculum on ever having had sexual intercourse at the end of program implementation?

C. Secondary research question(s)

Two secondary research questions were addressed:
- What is the impact of the TOP® relative to a WR curriculum on youth intentions to have sexual intercourse in the next year, intentions to use condoms if having sexual intercourse, and intentions to use effective means of birth control if having sexual intercourse in the next year at the end of program implementation?
- Is TOP® more effective for those deemed to be at higher sexual risk at baseline on both the primary outcome of ever having had sexual intercourse and on youth sexual intentions in the next year at the end of program implementation?

II. Program and comparison programming

TOP® is a youth development program designed to reduce teenage pregnancy and increase school success by helping youth develop a positive self-image, life management skills, and optimistic yet realistic expectations. The TOP® program model consists of three components

implemented over nine consecutive months by trained adult facilitators: (1) weekly group sessions, (2) community service learning (CSL), and (3) positive adult guidance and support. The intended program dosage for each TOP® participant is a minimum of 25 weekly sessions (40-50 minutes each) once per week and at least 20 hours of CSL over the nine months.

A. Description of program as intended

TOP® is a 9-month positive youth development program that provides structured small-group activities, caring adult support and guidance, and CSL in addition to basic sexuality education. The TOP® *Changing Scenes*® curriculum is separated into four age/stage-appropriate levels and includes topics such as goal setting, clarifying values, and decision-making. The curriculum is designed to facilitate the development of key individual life skills including decision-making, autonomy, and competence, and enhance competencies in interacting with adults and peers. It has been suggested that, consistent with other positive youth development programs, content regarding healthy sexuality and sexual risk reduction is enhanced by engagement in CSL activities, which impacts critical mediators for the youth such as autonomy, competence in decision-making, and interactions with adults, and recognition of positive future life options, resulting in motivation to not engage in risky sexual behaviors. Youth are engaged in all aspects of needs assessment, planning, and evaluation of their community service learning project. Each youth is expected to complete a minimum of 20 CSL hours.

TOP® facilitators' qualifications follow recommendations by Wyman and include (but are not limited to) knowledge related to adolescent and youth development; skills regarding communication, establishing trust and respect; being nonjudgmental; and having a positive attitude toward diversity, individual learning styles, shared decision-making power with youth, and youth voice. TOP® facilitators are all trained in the curriculum prior to delivering content. Each TOP® club is made up of 10 to 25 youth. Clubs meet weekly either after school or on weekends over a 9-month period; sessions typically last between 40 to 60 minutes in length. Clubs need to provide a minimum of 25 sessions over the 9-month time frame, but can have as many as 36. All sessions are implemented according to a program manual and each weekly session includes specific activities, such as brainstorming, games, and group discussions, to meet the session objectives.

B. Description of counterfactual condition

Youth in the counterfactual condition received a less intensive WR intervention; youth met with a facilitator once a month for 60 to 90 minutes to learn workplace competencies. Topics included building customer-relations skills, creating a job preparation portfolio, interview strategies, and appropriate dress and behavior for the workplace. WR sessions were held at the respective recreation centers and led by adults trained in the curriculum. In Year 1 of the program, adults from an outside community-based organization (CBO; different from the 4 CBOs providing TOP®) were hired to facilitate the WR sessions; in years 2 and 3 site-based employees facilitated sessions.

Early on, City of Rochester government felt that the control group youth needed to receive some organized programming from the study. However, it was clear that such programming needed to be relevant to the youth of Rochester, not focused on sexuality and/or other life skills

targeted by TOP®, and could be woven into the existing structure of the participating sites. The two most feasible alternatives concerned work readiness or healthy nutrition/eating. We chose work readiness for a number of reasons: the City of Rochester has the Summer of Opportunity Program (SOUP) that employs youth (aged 14-16) over the summer months and we felt that the WR curriculum could help ready youth for taking advantage of SOUP; the City of Rochester had an existing curriculum already developed that was easily modifiable for the time and age constraints of the current program; and the City of Rochester was, at the time, already conducting a large randomized trial of obesity prevention with similar aged youth.

III. Study design

A cluster-randomized design was used to estimate the impact of TOP® on reducing initiation of sexual activity among young (aged 11-14) urban teens in Rochester, New York. In all instances, impact analyses follow an intent-to-treat protocol, assessing the impact of offering the program to the youth. Given that random assignment, when implemented well, generally ensures that any baseline differences in group characteristics are the result of chance alone, differences in outcomes between the two groups can, thus, be causally attributed to the intervention alone. Given that researchers have little control over the participants within each cluster and that on average like individuals will tend to attend similar clusters (e.g., given that Hispanic youth tend to live in similar neighborhoods, they similarly are likely to attend similar recreation centers), cluster randomization has implications for baseline equivalence, particularly when the number of clusters is small. A mixed-method implementation study describes program implementation and provides context for the impact findings. The following sections describe in more detail sample recruitment and randomization, data collection methods, outcomes for the impact analyses, formation of the analytic sample and baseline equivalence of the study groups, and the analytic approach for both the impact and implementation studies. The full study protocol and all study procedures were approved by the University of Rochester's Office of Human Subjects Protection.

A. Sample recruitment

Cluster Recruitment. All school-based and free-standing City of Rochester Recreation Centers were reviewed by the City of Rochester Department of Recreation and Youth Services leadership to determine eligibility and feasibility for participation in this research project based on: a) location in zip codes with high teen pregnancy rates or serving youth who reside in such zip codes; b) attributes of the recreation centers themselves (physical size, available rooms, etc.); and c) attributes of youth attending recreation centers (participation rates in other recreation center activities, demographics of youth served, etc.). Using these criteria a total of 11 sites were chosen for project participation. Eight sites were free-standing recreation centers and three sites were out-of-school programs hosted in city schools. During the 3 years of the study, 3 of the sites were dropped during the summer months and were replaced. Because it was during the summer months, this did not affect study programming (which occurred from October to June of each year). Between year 1 and year 2, one was dropped as it was difficult to obtain appropriate space for the clubs at this site. Between year 2 and year 3, one site (the replacement site from the year 1 change) was dropped based on a mutual decision between the school principal and City project leadership to not continue at this school. Additionally, one of the smaller sites was dropped due to low attendance and a mutual decision by the recreation director and City project leadership

that recruitment to the project was likely to go better at a nearby school which was physically close to this center and which had developed more recreation programming with the City since the start of the THRIVE project, and the fact that most of the 11-14 year old youth from these neighborhoods would attend this particular site rather than the removed site, which attracted older teens. All sites were recruited into the study prior to randomization. Appendix A presents sample flow at the cluster level.

Each site was affiliated with one of four CBOs, which employed the TOP® facilitators. The four participating Rochester CBOs were: the YWCA working in two centers, Community Place of Greater Rochester working in two centers, Coordinated Care Services, Inc. working in three recreation centers, and Metro Council for Teen Potential working in four centers. Each CBO provides an extensive amount of youth programming in the City of Rochester, has expertise in the issues of youth sexuality and teen pregnancy, and has experience providing teen pregnancy/adolescent sexuality programs to city youth. Moreover, it was cost effective to employ CBO employees to deliver the program, rather than hire additional city employees.

Youth Recruitment. All youth 11 to 14 years of age enrolled at, living near, or attending school within the catchment area of one of the 11 randomly assigned sites as of September 2012 (Cohort 1), September 2013 (Cohort 2), and September 2014 (Cohort 3) were eligible for the study. All youth had to have a basic understanding of the English language (the TOP® curriculum has yet to be translated to a language other than English) in order to participate in the project. Youth were not able to participate in THRIVE (the study programming whether TOP® or WR) more than one year.

TOP® facilitators and City of Rochester Bureau of Recreation and Youth Services staff were responsible for recruitment efforts. Recruitment entailed a variety of informal and formal efforts conducted over the summer months with facilitators and Bureau of Recreation and Youth Services staff conducting formal presentations to parents and youth at recreation centers and in nearby or connected schools, canvassing of youth in the participating recreation centers to gauge potential interest, and canvassing of youth in neighborhoods. In all instances, parents/youth were screened to confirm age appropriateness, understanding of randomization, and to gauge barriers as well as willingness to participation. Evaluation team members were present at all formal presentations to explain the details of the research aspects of the project, answer any questions, and provide information about the study.

All interested youth/parents received an informational packet that included program and evaluation activities entailed in the study. Information was provided in each packet to address commonly asked questions, including both programming and evaluation team contact information. Within each packet was a recreation center registration form to formally register youth in the City Recreation system and to provide contact information through the RecPass system. The RecPass system was one strategy used by the research team to maintain contact with each participant.

Each participating site had target numbers for yearly enrollment to satisfy statistical power and other pragmatic considerations (including size of the recreation center). While each center would annually recruit an average of 40 youth to achieve the targeted 440 required to ensure adequate statistical power, recruiting 40 at each site was unrealistic due to size and numbers of

youth attending the smaller centers. For this reason, the targets presented in Table 1 have been adapted. Underlying these numbers was the reality that 20 youth needed to be recruited to each group to ensure that an average of 10 or more TOP® youth attended at least 75% of the sessions.

For youth/families that expressed interest and passed the informal vetting process, parental consent forms were distributed in late August of each year (all presentations conducted after consents had been distributed included copies of parent consent forms). If parent consents were not provided in person, consents were returned to the recreation center. Each returned consent was documented (youth name, date returned) and kept in a secure holding container within each study center (typically locked box in Recreation Center Director's office). Evaluation team members retrieved copies of tracking logs and signed consents at least twice weekly. Parental consent/permission and youth assent were both required for participation in the evaluation and in programming.

B. Study design

This is a cluster randomized trial conducted over three years, recruiting three cohorts of youth who received either TOP® or the WR counterfactual control. Randomization was conducted annually. Eleven sites were recruited and randomized each year for 3 years for a total of 33 randomized clusters (or out-of-school program offerings). Randomization was conducted by the evaluation team, using a random number generator. Each year the 11 out-of- school programs were randomized to condition, stratified by CBO providing TOP® programming. Two strata included two programs. One strata included three programs, with two centers being assigned to intervention and one to comparison each year. The fourth strata included four programs. Randomization occurred following the completion of baseline surveys, typically in late September.

Table 1. Annual targeted numbers of youth participants by CBO and site.

Community agency and total # of youth	Site #	# of groups of 20 youth	# of youth per Site
Agency A (n=80)	School 1[a]	1	20
	Recreation Center 2	2	40
	School 2	1	20
Agency B (n=80)	Recreation Center 4[b]	2	40
	Recreation Center 3	2	40
Agency C (n=180)	Recreation Center 4	2	40
	Recreation Center 5	2	40
	Recreation Center 6	3	60
	School 3	2	40
Agency D (n=100)	Recreation Center 7	2	40
	Recreation Center 8	3	60
Total	11 sites/year	22 groups/year	440 enrolled youth/year

[a] The free-standing center dropped after Year 2 and was replaced with school-based center.
[b] Dropped after Year 1 and replaced with school-based center. School-based center dropped after Year 2 and replaced with free-standing center.

C. Data collection

Impact evaluation data were collected primarily via youth paper and pencil surveys at two time points: baseline (prior to randomization) and immediate post intervention. There were three sections to the survey: Section A, which was completed by all youth and collected demographic information, most outcome data, and data on risk and protective factors. Sexual behavior items were collected using either Section B or Section C, depending upon whether the youth was sexually active. Table A.2 in Appendix A presents the timing of data collection efforts used in the impact analysis of TOP®.

Data on program fidelity and attendance, other teen pregnancy prevention programming in the area, and factors that may have affected program implementation were collected on an ongoing basis throughout the study period to document program implementation and provide context for the intent-to-treat impact findings.

1. Impact evaluation

All surveys were administered by members of the evaluation team. At each survey administration, youth were reminded of data confidentiality. Youth were also reminded that the survey was not a test, that there were no right or wrong answers, but that the information provided would be used to help improve services for other youth in Rochester. For each completed survey, youth were provided a $10 gift card in appreciation for their work. Though procedures changed a bit depending on data point, procedures were the same across experimental arms.

Baseline survey procedure. All baseline surveying was conducted prior to the start of program sessions (throughout September following Labor Day weekend). During this time, program facilitators met weekly with consented youth to introduce them to the study and to provide generic out-of-school activities to maintain engagement (e.g., cooking class, chess meet, and basketball tournament). These meetings occurred during the scheduled allotted program time. During these meeting times, baseline surveying was conducted in a group setting at the sites.

For each baseline survey, youth were required first to agree to participate and confirm this agreement by signing a youth assent form (youth who were not surveyed at baseline were still allowed to participate, but a signed assent form was required before any evaluation activities were conducted with the youth). Attendance at these sessions was aided by the facilitator at each site reminding participating youth of the need to complete baseline surveys. Facilitators often made reminder calls and/or provided transportation to assist with baseline survey completion. While the majority of participating youth were reached using these strategies, smaller groups were conducted for youth unable to attend the scheduled sessions by having the evaluation team work directly with the facilitator at each site. In some instances, phone surveys were conducted with the youth. In other instances, surveys were conducted in the home or at a mutually agreed upon site other than the study site (e.g., school, library).

Post-program survey procedure. Similar to the baseline survey procedure, the evaluation team focused on a one-month window (mid-May to mid-June) to obtain youth post-program

surveys immediately at the conclusion of the 9-month study, during regularly scheduled meeting times. At WR program offerings (which met monthly), the facilitator notified youth in attendance of the upcoming dates of surveys. For the WR control arm, which met only monthly, evaluation team members were at each recreation center site on the day and time of the week that the group met for consecutive weeks throughout the monthly window (e.g., if a WR group met on the third Thursday of the month at 4:30pm, the evaluation team was at that recreation center each Thursday at 4:30pm to survey youth). At TOP® sites (which met weekly), the TOP® facilitator notified each youth of the upcoming survey, worked with the evaluation team to facilitate logistics (space within each center), and, in subsequent weeks worked with the youth who had already been surveyed (provided similar generic activities as described above [non-TOP® programming]). Evaluation team members worked with the WR facilitator and recreation center staff to facilitate survey logistics. Incentives for completion of the post-program survey included the $10 gift card as well as either a field trip to the local Six Flags of America or a pass to a smaller local amusement park that the youth could use throughout the summer months.

For youth not surveyed within the month of mid-May to mid-June, the evaluation team worked with TOP® and WR facilitators and other Bureau of Youth Services and Recreation staff to contact and schedule youth for post-program surveying. These surveys were conducted either individually or in small groups at a site convenient to the youth, with an evaluation team member(s) present. While most post-program surveys were collected early in the summer months, we continued with this approach throughout July in order to obtain as many surveys as possible

2. Implementation evaluation

A *fidelity to implementation plan* was developed in July of 2011 and was structured to address fidelity in regard to content, pedagogy, and implementation. Table B.1 in Appendix B summarizes the data sources used to assess the core implementation elements, including the frequency of data collection and the staff responsible for collection. The implementation evaluation included an assessment of adherence to the *Changing Scenes®* curriculum and TOP® programming, the quality of implementation, experiences of the participants in the counterfactual program, and context.

Adherence and quality of implementation were measured by fidelity monitoring logs (FMLs) that were developed in collaboration with a team of TOP® grantees based on information provided by Wyman and knowledge of appropriate fidelity assessment. The FML was completed immediately following the session using SurveyMonkey. FMLs included name of facilitator, date of session, length of session, number of participants, adherence to activities being provided as per the curriculum, as well as other information regarding pedagogy and implementation. Each FML was reviewed by an evaluation team member on a weekly basis, with feedback being provided to the Project Director. Any FML items that were identified as not following the *Changing Scenes®* curriculum were reviewed with the TOP® facilitator and that person's supervisor by the Project Director. Careful attention was paid to the possibility of "interventionist drift." Attendance data were collected using an Excel spreadsheet.

Facilitators of the WR programming also completed FMLs demonstrating that they were not providing content related to sex, sexuality, or pregnancy prevention. These FMLs were

developed by the evaluation team and were completed in SurveyMonkey within 24 hours of a session.

Ten percent of all TOP® and WR sessions were also observed by an outside observer separate from program staff. All observers were trained in the *Changing Scenes*® curriculum by a Wyman credentialed trainer prior to observing sessions. Inter-rater reliability of observers was conducted in 2013 (10 sessions) and found to be acceptable (weighted kappas were all above .80 for overall program quality, ability to demonstrate a "values neutral" approach, and an ability to effectively address questions and intervene in conflict; above .70 for knowledge of the program and level of enthusiasm; below .70 for rapport with the youth). Observation forms and FMLs were completed by each observer within 24 hours of observing a session and were reviewed weekly by the evaluation team. Any areas of concern identified by the evaluation team were reviewed with the Project Director.

D. Outcomes for impact analyses

The primary research question was answered using a single-item dichotomous measure from the immediate post-intervention survey: "*Have you ever had sexual intercourse?*" This measure of ever having had sexual intercourse captured the effect offering TOP® on the full sample in delaying the onset of sexual intercourse (see Table 2).

The secondary research questions concerned intentions to have sexual intercourse in the next year and intentions to use sexual protection in the next year and were assessed with the following three questions, answered by all participating youth, irrespective of whether or not they were sexually active:

- *Do you intend to have sexual intercourse in the next year?*
- *If you have sexual intercourse in the next year, do you intend to use (or have your partner use) a condom?*
- *If you have sexual intercourse in the next year, do you intend to use an effective method of birth control (for example, condoms, birth control pills, the shot, the patch, the ring, IUD, implant)?[1]*

To address the secondary research question regarding risk status at baseline, two questions were used to define sexual risk at baseline: ever having had sexual intercourse and the intention to have sexual intercourse in the next year. The moderation analyses are discussed in more detail below.

E. Study sample

Table C.1. in Appendix C depicts the flow of sample members from the beginning of the study through the post-intervention assessment point. Sixty percent of the 1,978 youth who

---

[1] The sexual intentions items were only asked of sexually active youth at baseline of Year 1 (cohort 1). Missing baseline intention data for cohort 1 youth at baseline was imputed using expectation-maximization imputation following the procedures outlined by Allison (2002) with other baseline covariates included as auxiliary variables, following the same strategies as the multiple imputation approach discussed in Appendix G.

Table 2. Behavioral outcomes used for primary impact analysis research question.

Primary impact analysis outcome name	Description of outcome	Timing of measure relative to program
Ever having had sexual intercourse	*This variable was a yes/no measure of whether a youth had ever had sexual intercourse. The measure was asked of all participating youth and taken directly from the following item on the survey:* • *"Have you ever had sexual intercourse?"* *The variable was constructed as a dummy variable with "yes" coded as 1, "no" coded as 0, and coded missing if left unanswered.*	*Immediate post intervention*
Secondary impact analysis outcome names		
Intention to have sexual intercourse [a]	*This variable was a single item measure of each youth's intentions to have sexual intercourse in the next year. The measure was asked of all participating youth and taken directly from the following item on the survey:* • *"Do you intend to have sexual intercourse in the next year?"* *The variable was constructed based on youth's response using a 4-point Likert scale (4 = Yes, definitely; 3 = Yes, probably; 2 = No probably not; 1 = No, definitely not). Unanswered responses were coded as missing.*	*Immediate post intervention*
Intention to use (or have partner use) condoms [a]	*This variable was a single item measure of each youth's intentions to use condoms (or have partner use) if having sexual intercourse in the next year. The measure was asked of all youth and taken directly from the following item on the survey:* • *"If you have sexual intercourse in the next year, do you intend to use (or have your partner use) a condom?"* *The variable was constructed based on youth's response using a 4-point Likert scale (4 = Yes, definitely; 3 = Yes, probably; 2 = No probably not; 1 = No, definitely not). Unanswered responses are coded as missing.*	*Immediate post intervention*
Intention to use effective means of birth control [a]	*This variable was a single item measure of each youth's intentions to use an effective means of birth control if having sexual intercourse in the next year. The measure was asked of all youth and taken directly from the following item on the survey:* • *"If you have sexual intercourse in the next year, do you intend to use an effective method of birth control (for example, condoms, birth control pills, the shot, the patch, the ring, IUD, implant)?"* *The variable was constructed based on youth's response using a 4-point Likert scale (4 = Yes, definitely; 3 = Yes, probably; 2 = No probably not; 1 = No, definitely not). Unanswered responses are coded as missing.*	*Immediate post intervention*

Source: Helping Youth THRIVE: Youth Development Survey administered at immediate post-intervention.

[a] Only asked of sexually active youth at baseline of Year 1 (cohort 1). Missing baseline intention data for cohort 1 youth at baseline was imputed using expectation-maximization imputation.

expressed interest provided parental consent and were eligible for participation in the study (n = 1,188). Out of these eligible sample members, 86% (n = 1,018) completed the baseline survey (treatment group = 571, 84%; control group = 447, 88%). At immediate post intervention, 81% of the youth completed post surveys (treatment group = 562, 83%; control group = 399, 78%). The attrition rate at immediate post intervention was 19%, with differential attrition of 4.9 percentage points. Seventy-three percent of youth completed both the baseline and post-intervention surveys. Only 80 youth (7%) provided neither a baseline or post intervention survey; 146 (12%) provided baseline information only; and 90 (8%) provided post intervention information only. For a youth to be included in the analytic sample, they must have had provided data on all outcome measures (at both baseline and post) and their gender, age, ethnicity, and race. If not collected at baseline, demographic data was imputed from data collected at other waves of the study. With these restrictions, 708 youth comprised the analytic sample (60% of original consented sample; 414 TOP®, 61%, 294 control, 58%).

In general, youth in the analytic sample were in early adolescence, racially and ethnically diverse, and not engaging in sexual risk-taking behavior at baseline. A little over one half (54%) were female, with an average age of 12.4 years. The majority reported being Black/African-American only (65%); 15% reported being White/Caucasian only; 9% reported being both Black/African-American and White/Caucasian; 6% reported being Black/African-American and other race (Asian, Hawaiian/Pacific Islander, Native American/Alaskan Native, or other race), 3% reported being Black/African-American, White/Caucasian and other race, and 2% reported being of other race only. Eighteen percent reported being multi-racial. Thirty-one percent reported being of Hispanic ethnicity. The majority reported speaking English at home (96%); 16% also reported speaking Spanish at home. Ninety-four percent had never had sex at baseline; 97% had not had sex "recently" with recently defined as the three months prior to the baseline survey. Of the 480 not in the final analytic sample, 80 had neither a baseline or post survey, 90 had a post survey only, 147 had a baseline survey and no post survey, and 163 had missing data on one or more of the demographic, baseline covariates, or outcomes of interest. Appendix D presents baseline differences between the analytic and non-analytic sample. Perhaps not surprisingly, youth not in the analytic sample were at somewhat higher risk at baseline, being more likely to be older, be males, and report more intentions to have sex in the next year (Ellickson, Bianca, & Schoeff, 1988; Greene, Lee, Constance, & Hynes, 2013; Weisman & Gottfredson, 2001).

F. Baseline equivalence

Despite the expectation that randomization would produce equivalent groups on all measured and unmeasured variables, we conducted baseline equivalence tests for demographics and baseline sexual outcomes to assess whether attrition affected the comparability of the treatment and control arms in the analytic sample. The statistical models for assessing baseline equivalence have the same structural form as the models used to estimate impacts (multilevel logistic and continuous regressions). Specifically, we tested for treatment versus control differences on the baseline value of each outcome variable for the primary and secondary research questions, as well as for the demographic variables of age, sex, and race/ethnicity. We used a multilevel model to account for the clustering of youth within site with experimental condition at the level of randomization (i.e., site) with cohort and stratification dummy variables included as covariates.

Table 3 summarizes the key baseline measures for the analytic sample. In terms of demographics, there were no age, gender, or race differences. There was, however, significant imbalance in the distribution of Hispanic youth, with the TOP® group comprised of more Hispanic youth. Finally, there were no differences in the youth sexual activity intentions in the next year or having had sex at baseline across the groups.[2]

G. Methods

To answer the primary research question, we used an intent-to-treat (ITT) framework (Fisher et al., 1990; Ten Have et al., 2008) and data collected at the immediate post-intervention assessment point to estimate the average impact of TOP®, relative to the WR control group, on participants ever having had sexual intercourse, and sexual intentions in the next year. An ITT analysis estimates the impact of the program on all eligible youth who were enrolled in a study site regardless of the level of program participation (TOP® or WR).

1. Impact evaluation

To address the question of TOP® impacts on the primary and secondary outcomes, the analytic approach used multilevel regression modeling to account for randomization at the site level as well as the clustering of youth within site. For the dichotomous outcome of ever having had sexual intercourse, a two-level logistic regression model was tested. The sexual intentions in the next year items were analyzed as continuously distributed using robust standard errors to adjust for non-normality in the distribution. To aid in interpretation and to limit problems associated with cross level correlations, all predictors were grand mean centered (Hofmann & Gavin, 1998). To improve precision and statistical power, both site (cohort, CBO stratification) and youth level baseline characteristics (gender, age, race/ethnicity, and baseline value of the dependent variable) were entered as covariates in all models tested. Impact estimates with p-values less than .05 (two-tailed tests) were considered statistically significant; providing evidence that there are likely true differences between the groups as a result of TOP®. Detailed model specifications for baseline differences are provided in Appendix E with detailed model specifications for the impact analysis presented in Appendix F.

To assess the secondary outcome associated with baseline sexual risk of the youth, a cross-level interaction term was entered for each primary and secondary outcome. As previously alluded, two separate risk interaction analyses were run. The first defined risk as those who reported being sexually active at baseline, with the treatment condition by having had sex at baseline comprising this term. This interaction term was entered only for the secondary research question of sexual intentions. The second defined risk by the response to the intention to have sexual intercourse in the next year question at baseline, with treatment condition by intention to have sex forming this term. In both instances, we hypothesized that TOP® would be more effective with those at higher baseline sexual risk. If significant interaction effects are detected, results will be graphically examined following Aiken and West (1991).

---

[2] There were, however, marginal differences in gender (p = .070), other race only (p = .079), and having had sexual intercourse at baseline (p = .095).

Missing data occurred at both baseline and immediate-post intervention data collection points. Missing baseline values have practical implications as most multilevel software invokes

Table 3. Summary statistics of key baseline measures for youth completing both baseline and immediate post intervention survey.

Baseline measure	n	Intervention Mean or % (standard deviation)	n	Comparison Mean or % (standard deviation)	Intervention versus comparison mean difference	Intervention versus comparison p-value of difference
Demographics						
Age	414	12.26 (1.14)	294	12.48 (1.09)	-0.22	.222
Gender (female)	414	57.2%	294	48.6%	8.6%	.070
Ethnicity: Hispanic	414	36.5%	294	24.1%	12.4%	.007
Race: White only	414	16.7%	294	13.6%	3.1%	.851
Race: Black/African-American only	414	61.6%	294	68.7%	-7.1%	.494
Race: Black/African-American and White	414	9.2%	294	7.8%	1.4%	.757
Race: Black/African-American and other race(s)	414	5.8%	294	5.4%	0.4%	.964
Race: Black/African-American and White and other race(s)	414	3.4%	294	3.4%	0.0%	.985
Race: Other race only (Reference)	414	3.4%	294	1.0%	2.4%	.079
Sexual Outcome						
Ever having had sexual intercourse	414	5.1%	294	8.5%	-3.4%	.095
Sexual Intentions Outcomes						
Intention to have sexual intercourse	414	1.93 (0.95)	294	1.88 (0.99)	0.05	.586
Intention to use (or have partner use) condoms	414	2.74 (1.18)	294	2.73 (1.26)	0.01	.894
Intention to use effective means of birth control	414	2.67 (1.15)	294	2.68 (1.24)	-0.01	.965

Source: Helping Youth THRIVE: Youth Development Survey administered at baseline assessment.,

Note: Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.

listwise deletion of missing data by default (sample sizes ranged from 89-127 missing demographic and/or baseline value of the dependent variable). Four types of sensitivity analyses

were conducted. The first used an ITT approach but imputed missing baseline data. The second focused on matched sample that was identified using propensity score matching. The third tested delay of onset of having sex for youth who were sexually naïve at baseline and the fourth sensitivity analysis treated inconsistent sexual behavior responses as missing (youth who reported ever having had sexual intercourse at baseline and who reported not ever having had sexual intercourse at post intervention). Appendix G provides results from these sensitivity analyses.

For the sexual activity question, inconsistent responses were possible (e.g., youth reports never having had sexual intercourse at post-intervention but reported being sexually active at baseline). Appendix G, section 4 presents information on the amounts of inconsistent responses for each outcome and results analyzing the raw data. In the main analytic sample, inconsistent responses within survey were coded as never having had sexual intercourse for a variety of reasons (youth would report having had sexual intercourse in Part A of the survey and report not having had sexual intercourse and completing the survey section devoted to non-sexually active youth (Part C); youth wrote that he/she had never had sexual intercourse when completing the survey section devoted to sexually active youth; youth would give age of timing of first sexual encounter as being far into the future) while the raw data were analyzed for the two youth who reported ever having had sexual intercourse at baseline and reported being sexually naïve at post intervention assessment.

2. Implementation evaluation

The implementation study focused on four areas: the extent to which the program adhered to program fidelity standards and was delivered with quality, as well as the experiences of the control group and any contextual circumstances that substantially affected implementation.

Data sources included Fidelity Monitoring Logs (FML) from facilitators and observers for both the intervention and counterfactual conditions, observation forms completed by the observers for the intervention and counterfactual conditions, and attendance forms which were reviewed and analyzed to ascertain adherence, quality, and context. In regard to adherence and quality, analyses included descriptive statistics including calculating numbers and percentages, as well as means and medians. In regard to context, reports maintained by the Project Director were reviewed and findings were summarized.

Appendix H presents further information regarding implementation evaluation methods.

IV. Study findings

The two goals of the evaluation were to: (1) determine if TOP® had favorable impacts on younger adolescent sexual behavior/intentions and (2) understand how TOP® was implemented in this non-traditional TOP® setting to provide context for the impact findings. The next section presents the results of the implementation study, followed by findings from the impact analyses to determine the overall effectiveness of the intervention for younger adolescents.

A. Implementation study findings

TOP® Adherence. At the Rochester site, one modification to the TOP® model was incorporated with permission of both the Office of Adolescent Health and the publishers of the TOP curriculum, Wyman Center. The modification allowed for an assistant to be present at all TOP® club events/meetings (the original curriculum allows for an assistant for clubs with 25 or more youth), thus two adults were present at most TOP® club meetings. In Rochester, eleven adult facilitators provided TOP® content over the grant period. All facilitators met or exceeded position requirements. All facilitators were trained in the TOP® curriculum prior to leading sessions and received annual updates regarding the curriculum, the fidelity process, and the grant logic model. The offerings by program staff in this instance were generally consistent with the TOP® model: Youth were offered a minimum of 25 weekly sessions with an average of 27.6 sessions. Sessions were generally held weekly, with exceptions including when the sites were closed, when the facilitator was ill, when there was not a sufficient number of youth available to conduct a session, and over the holidays.

The average class period length was 60 minutes, which is 16-33% longer than expected. Recommended session duration suggested by Wyman varied depending on the level and content. Facilitators often mentioned that a session was "too short" for all the content to be provided, and indeed, we did have a few sessions where all activities were not completed. Activities were completed as per the *Changing Scenes®* curriculum 95.6% of the time, as per the facilitators self-report and 99.1% of the time according to observational data. All TOP® classes had the minimum ratio and number of trained facilitators, with an average student to staff ratio of 12:1. No adaptations were made to the curriculum over the course of the program.

The dosage received by treatment group members did not consistently meet program model expectations and varied substantially by site and by year. Appendix I presents TOP® attendance by site by year. Treatment group members attended a median of 18 weekly sessions, with 27% meeting or exceeding the minimum dosage of 25 sessions. Forty-eight percent of the treatment group members, however, attended 19 or more of the sessions (i.e., 75% or more). The median number of CSL hours completed by the treatment group was 13, with 36% completing the minimum 20 hours. The percentage of treatment group members who attended at least 25 sessions and completed a minimum of 20 CSL hours was 22%; 32% attended at least 19 sessions (75%) and completed a minimum of 20 CSL hours. Not surprisingly, session attendance was associated with the number of CSL hours completed ($r = .87$, $p < .001$). Importantly, the number of treatment group youth who failed to attend any TOP® sessions improved as facilitators became more aware of the importance of vetting potential families and youth to make sure that youth were committed to participation after year 1 (number of youth not attending a TOP® session; year 1 = 24%, year 2 = 10%, year 3 = 9%).

TOP® Quality. The *Changing Scenes®* curriculum was provided with high quality as noted by facilitators and observers alike. There was greater than 90% agreement with 5 of the 8 items measured, and greater than 80% agreement with the remaining 3 measures (see Table 4). Further, trained observers rated the facilitators as either 4 or 5 (scale of 1 to 5, with 5 being most positive) an average of 90% of the time on a number of constructs important to TOP® session quality (see Table 5). In instances where a facilitator received a 3 or less on any item, the program director met with that facilitator to address the concern.

Experiences of the Control Group. The FMLs completed by the WR facilitators suggested that the pedagogy and content of the counterfactual condition had some overlap with the TOP® intervention from a positive youth development perspective, but only minimal from a sexuality

Table 4. Facilitator self-report and observer report of TOP® sessions.

	Facilitator % Yes	Observer % Yes
Was the facilitator adequately prepared for the lesson?	99.1	96.3
Did the facilitator do a "self-check" before the lesson?	99.1	98.8
Did the facilitator have all the supplies needed for the lesson?	98.9	95.1
Did the facilitator listen more than they talked?	84.8	86.4
Did the facilitator acknowledge and reward desirable behavior?	94.5	88.9
Did the facilitator elicit questions/responses from multiple members of the group?	97.4	97.5
Did youth participate in setting limits and rules?	86.8	92.6
Did the facilitator employ ELC techniques while facilitating the lesson?	82.8	97.5

Source: Facilitator completed Fidelity Monitoring Log (FML)
Observer completed Fidelity Monitoring Log (FML)

Note: ELC = Experiential Learning Cycle

perspective (see Table 6). A small percentage of sessions had FMLs completed (32% of possible sessions). Participation in the WR programming was poor: only a small amount (< 10%) of participants were documented to have received content.

Table 5. Observer ratings for each 12 constructs assessing TOP® program quality (n = 96).

	Rating of 4 or more* (%)
In general, how clear were the program implementer's explanations of activities?	91.7
To what extent did the implementer keep track of time during the session and activities?	88.5
To what extent did the presentation of materials seem rushed or hurried?	91.7
To what extent did the participants appear to understand the material?	87.6
How actively did the group members participate in discussions and activities?	93.8
Facilitator's knowledge of the program	95.9
Facilitator's level of enthusiasm	84.4
Facilitator's poise and confidence	88.6
Facilitator's rapport and communication with participants	87.6
Facilitator's ability to effectively address questions/concerns	88.6
Facilitator's ability to effectively intervene to address any conflicts within the group in a respectful and supportive manner	91.7

Facilitator's ability to demonstrate a "values neutral" approach through the lesson/activity	93.7

Note: The survey used a 5-point scale in which 5 was the most positive response.

Table 6. Summary fidelity assessment information for Work Readiness (counterfactual) condition.

	Facilitator FML's for 93 sessions	Observer 10 sessions observed
Median number of participants	8	8
% decision making content	61.4%	62.5%
% goal setting content	56.6%	62.5%
% communication/assertiveness content	54.2%	62.5%
% values setting content	24.0%	83.3%
% romance content	2.4%	12.5%
% sexuality content	2.4%	12.5%
% contraception content	0.0%	0.0%
% pregnancy content	0.0%	0.0%
% service learning content	18.1%	33.3%
% group building activities content	49.4%	75.0%
% service learning activities	19.3%	20.0%

Note: Any amount of the construct offered/observed was counted as present. The amount offered during the session was not captured.

Context. The City of Rochester has a variety of agencies providing sexual education/prevention or HIV prevention programming in a variety of settings. Each year of the post survey, we asked youth if they had participated in TOP® or WR at their site. For the WR youth, 15.2% reported participating in the TOP®; 24.3% reported participating in a WR program. For the TOP® youth, 28.8% reported participating in TOP®; 15.9% reported participating in a WR program. It is important to point out, however, that both facilitators and youth may not have self-identified their club with the TOP® name as the overall study was better known locally as THRIVE and each club self-identified a club name. Additionally, we asked youth if they had participated in a work or job readiness program outside of the site during the year; 9% reported participating in a work or job readiness program outside of the site (11.8% of the WR youth; 6.1% of the TOP® youth). Similarly, we asked youth if they had participated in any sexual education/prevention or HIV prevention programs being offered throughout Rochester. We were aware of seven local programs being offered (including TOP® [3]) and asked youth about participation in each program by name. We also asked if they participated in any other sexual education/prevention or HIV prevention program. In total, 18.4% of the youth reported participating in one or more of these programs (19.2% of WR youth and 17.9% of TOP® youth). Additionally, since all sixth graders in the city school district are required to take a health class which teaches some aspects of sexual education, the analytic sample likely received some sexual education before or during their participation in our programming (though a number of youth attended charter schools in Rochester and we were unaware of the level of systematic sexual health education provided in city charter schools).

_____

[3] Aside from the study, TOP® was being used by one local child service agency with a focus more on high-risk, seriously emotionally disturbed youth.

In terms of contextual factors influencing implementation, weather conditions (snow & cold temperatures), competition with other recreational/sports activities, physical space in some sites, and change in Rochester City School District schedule (extended day) each affected program implementation. For instance, in the winter of Year 2 of the program, Rochester suffered extreme cold for over a week period. During this time, recreation centers were used for emergency shelters for the homeless and programming could not be offered. Over the 3 years, Rochester schools also had a number of snow days which also affected programming. While the sites offered flexibility in rescheduling these missed days, the make-up days were often not attended as well as the usual meeting time as youth/families had other commitments on these make up days or communication regarding the change was unclear or indirect. In terms of recreational/sports activities, a number of youth participated in structured activities or sports (center based and/or school based) that conflicted with programming. Despite vetting youth and families at the start of programming, many did not consider scheduling that far in advance. While youth were allowed to miss chunks of programming (if necessary) during a season and then come back to the club, these seasonal activities affected attendance. Also in Year 2 of the program, the local school district enabled many schools to move to an extended day schedule which had the school day going until 5pm or later for many youth. Even though programming was adjusted in many sites, this had an effect in at least one center where the program could no longer continue as the center opted for other services to take priority in this compressed time.

B. Impact study findings

Primary Research Question. There is no evidence that TOP® caused changes in the likelihood of initiating sexual activity. At the immediate-post intervention data collection point, about 11% of treatment group members reported ever having had sexual intercourse, compared to 16% of the control group. The estimated impact is not statistically significant ($p = .610$), once the baseline demographic differences between groups is accounted for, indicating there is likely no true difference between the two groups. Table 7 presents the estimated effects of TOP® on the primary outcome measure of ever having had sexual intercourse. As presented in Appendix G, all sensitivity analyses corroborate this finding.

Secondary Research Questions. There is also no evidence that TOP® had an effect on any of the three sexual intentions assessed (Table 8). Most youth reported that they were not likely to have sexual intercourse in the next year, but if they did, were likely to use a condom and/or an effective means of birth control. In terms of TOP® effects on high-risk sexual youth, Table 3

Table 7. Post-intervention estimated effects for the primary research question of ever having had sexual intercourse.

Outcome measure	Intervention Sample Size (N)	Intervention %	Comparison Sample Size (N)	Comparison %	Intervention compared to comparison mean difference (*p*-value)
Ever having had sexual intercourse	414	11.4%	294	16.0%	-4.6% (.610)

Source: Helping Youth THRIVE: Youth Development Survey administered at immediate post-intervention.

Analyses statistically control for sex, age, ethnicity/race, and baseline value of the dependent variable at the youth level and cohort and CBO stratification variables at the recreation center level. See Table 2 for a more detailed description of each measure and section III.6.1 for a description of the impact estimation methods.

presents baseline rates of sexual activity by group and mean scores on the intentions to have sexual intercourse in the next year item. A specific breakdown of the baseline sexual intentions item indicates that 5% of the youth responded "Yes, definitely" (5% TOP®; 6% WR), 27% responded "Yes, probably" (28% TOP®; 25% WR), 22% responded "No, probably not" (23% TOP®; 19% WR), and 46% responded "No, definitely not" (44% TOP®; 49% WR). Again, sensitivity analyses presented in Appendix G corroborate these findings.

In terms of the moderation effect of baseline sexual intentions by treatment condition on the primary outcome of ever having had sexual intercourse at post or the moderation effect of baseline ever having had sex by treatment condition on the secondary outcomes of sexual intentions, there was no evidence of moderation. However, in terms of the moderation effects of baseline sexual intentions to have sexual intercourse in the next year on post intentions to have sexual intercourse, there was a significant interaction effect ($p = .001$). Figure 1 graphically illustrates this finding. As shown in the figure, TOP® appears to have a beneficial impact on intentions to have sexual intercourse for the group of youth who reported that they were likely to have sexual intercourse in the next year at baseline ("Yes, definitely" and "Yes, probably" groups), with outcomes more comparable for those youth who not likely to have sexual intercourse in the next year at baseline ("No, definitely not" and "No, probably not" groups). To the extent that intentions translate to behavior, this finding merits more attention with further longitudinal data on sexual behaviors (as mentioned, this interaction effect was not present for ever having had sex at post-test). There were no significant interaction effects on sexual intentions to use condoms or sexual intentions to use an effective method of birth control. Sensitivity findings for these moderational analyses, however, were mixed. Imputation of missing baseline covariates corroborated these findings, including the significant condition by sexual intentions interaction effect on intentions to have sexual intercourse in the next year. Propensity score analyses supported the null findings regarding having had sexual intercourse and intentions to use a condom and intentions to use an effective method of birth control in the next year. Propensity score analysis did not corroborate the significant moderational effect of condition by sexual intentions on intentions to have sexual intercourse in the next year. Here, results were not statistically significant (p = .217; see Appendix G).

Table 8. Post-intervention estimated effects: secondary research questions.

Outcome measure	Intervention Sample Size (N)	Intervention mean (standard deviation)	Comparison Sample Size (N)	Comparison mean (standard deviation)	Intervention compared to comparison mean difference (p-value)
Intention to have sexual intercourse	414	1.60 (0.94)	294	1.64 (1.01)	-.04 (.496)
Intention to use (or have partner use) condoms	414	3.19 (1.25)	294	3.10 (1.31)	.09 (.687)
Intention to use effective means of birth control	414	3.09 (1.27)	294	3.05 (1.32)	.04 (.825)

Source: Helping Youth THRIVE: Youth Development Survey administered at immediate post-intervention.

Notes: Analyses statistically control for sex, age, ethnicity/race, and baseline value of the dependent variable at the youth level and cohort and community service agency stratification variables at the recreation center level. See Table 2 for a more detailed description of each measure and section III.6.1. for a description of the impact estimation methods. Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.
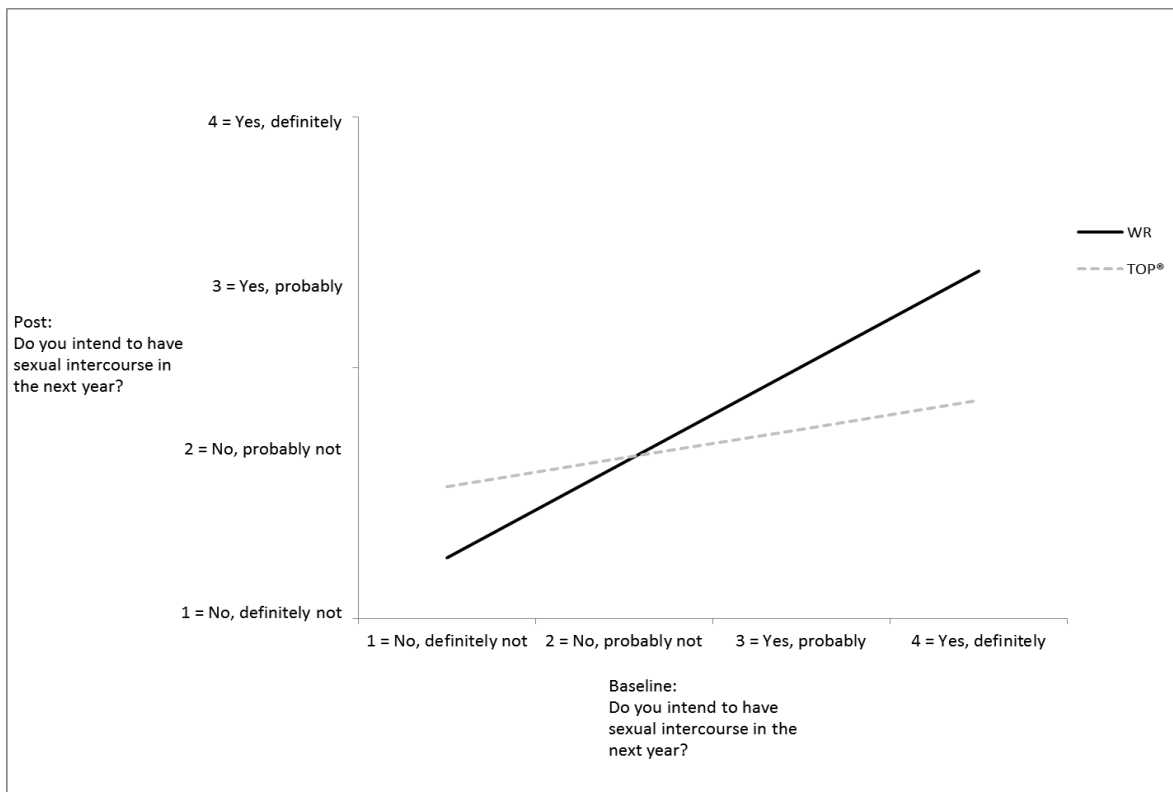

Figure 1. Baseline intentions to have sexual intercourse by condition interaction on post intentions to have sexual intercourse.

V. Conclusion

TOP® is one of the most widely used teen pregnancy prevention programs in the country (Wyman National Network, 2015). Compared with previous evaluations, both the setting and target age group were unique in the current trial. We implemented TOP® in urban recreation centers, a non-traditional setting for TOP®. We also focused on TOP® effects with younger adolescents (aged 11-14 at study entry) as the previous evidence was with samples comprised mostly of high-school aged older adolescents with TOP® being conducted in school settings (Allen & Philliber, 2001; Allen et al., 1997). In Rochester, the program was generally delivered as intended; however, many youth did not receive the minimum dosage of TOP® (22% attended 25 or more sessions and completed 20 or more hours of CSL; 32% attended 19 [75%] or more sessions and completed 20 or more hours of CSL). Based on data from over 800 youth from 11 sites in Rochester, New York, there was no impact on the behavioral sexual outcome of ever having had sexual intercourse. That is, TOP® youth were no less likely to report ever having had sex at the immediate post-intervention assessment point. Further, there were no differences between the groups on intention to have sex in the next year or on intentions to use a condom or an effective method of birth control if he/she did have sex in the next year.

In terms of the outcomes assessed, the younger age of our sample may have a bearing on these findings. That is, only a little more than 6% of our youth reported having had sexual intercourse at baseline. Our younger sample had a low rate of sexual activity and it appears that a longer term follow up may be necessary to detect differences in risky sexual behaviors for these youth. Though not presented in this report, future research will examine these outcomes in longer term (6 and 12-month) follow-up assessments.

There was an intriguing moderation effect on post assessment intentions to have sexual intercourse in the next year. Here, TOP® appears to have a stronger effect for those youth who reported having some intention to have sex in the next year (either definitely or probably) at baseline with effects more comparable among the larger group of youth who reported not having such intentions. This moderation effect, however, did not translate to sexual behavior (ever having had sex) at post. Future attention will be devoted to more longitudinal assessment of sexual behavior to examine the relationship among intentions and actual behaviors. Additionally, it may also be that this effect is even more pronounced among youth who were sexually naïve at baseline.

The current findings lead to three areas of future analyses that will be explored in additional publications. First, other moderation effects will be examined to assess if TOP® effects are more beneficial for youth who are at higher risk in other domains of life (academically and socioemotionally) prior to program entry. Second, given the variability in attendance and CSL completion we plan to undertake a complier average causal effect approach (Jo, Asparouhov, Muthen, Ialongo, & Brown, 2008; Jo & Muthen, 2001; Schochet & Chiang, 2011), examining program effects for those who do comply (attend sufficiently) to TOP® adherence fidelity. A third area for future analyses is to better understand attendance patterns at such youth programming and to better understand why youth did and did not attend TOP® programming throughout the nine month period.

24

VI. References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.

Allen, J. P., & Philliber, S. (2001). Who benefits most from a broadly targeted prevention program? Differential efficacy across populations in the Teen Outreach Program. *Journal of Community Psychology, 29*(6), 637-655.

Allen, J. P., Philliber, S., Herrling, S., & Kuperminc, G. P. (1997). Preventing teen pregnancy and academic failure: Experimental evaluation of a developmentally based approach. *Child Development, 68*(4), 729-742.

Allen, J. P., Philliber, S., & Hoggson, N. (1990). School-based prevention of teen-age pregnancy and school dropout: Process evaluation of the national replication of the Teen Outreach Program. *American Journal of Community Psychology, 18*(4), 505-524.

Allison, P. D. (2002). *Missing data*. Thousand Oaks, California: SAGE Publications, Inc.

Austin, P. C. (2010). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*, 150-161.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology, 163*(12), 1149-1156.

Cohen, J., & Cohen, P. (1985). *Applied multiple regression and correlation analysis for the behavioral sciences* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum and Associates.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351.

Coyne, C. A., & D'Onofrio, B. M. (2012). Some (but not much) progress toward understanding teenage childbearing: A review of research from the past decade. In J. B. Benson (Ed.), *Advances in Child Development and Behavior* (Vol. 42, pp. 113-152). San Diego, California: Elsevier Academic Press.

Doherty, E. J. (2015). *Benchmarking Rochester's poverty: A 2015 update and deeper analysis of poverty in the City of Rochster*. Retrieved from Rochester, New York: Rochester Area Community Foundation and ACT Rochester website: http://www.actrochester.org/sites/default/files/Poverty%20Report%20Update%202015-0108.pdf

Ellickson, P. L., Bianca, D., & Schoeff, D. C. (1988). Containing attrition in school-based research: An innovative approach. *Evaluation Review, 12*(4), 331-351.

Fisher, L. D., Dixon, D. O., Herson, J., Frankowski, R. K., Hearron, M. S., & Peace, K. E. (1990). Intention to treat in clinical trials. In K. E. Peace (Ed.), *Statistical issues in drug research and development (American Statistical Associations Group)* (pp. 331-350). New York: Marcel Dekker.

Greene, K. M., Lee, B., Constance, N., & Hynes, K. (2013). Examining youth and program predictors of engagement in out-of-school time programs. *Journal of Youth and Adolescence, 42*(10), 1557-1572.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods, 15*(3), 234-249.

Hoffman, S. D., & Maynard, R. A. (2008). *Kids having kids: Economic costs and social consequences of teen pregnancy* (2nd Ed.). Washington, DC: Urban Institute Press.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*(5), 623-641.

Jo, B., Asparouhov, T., Muthen, B. O., Ialongo, N. S., & Brown, C. H. (2008). Cluster randomized trials with treatment noncompliance. *Psychological Methods, 13*(1), 1-18.

Jo, B., & Muthen, B. O. (2001). Modeling of intervention effects with noncompliance: A latent variable approach for randomized trials. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 57-87). Mahwah, New Jersey: Lawrence Erlbaum and Associates.

Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Matthews, T. J., & Osterman, M. J. K. (2010). *Births: Final Data for 2008*. Retrieved from Hyattsville, Maryland: Center for Disease Control and Prevention website: http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_01.pdf

Metro Council for Teen Potential. (2012). Teen Birth Rates, Ages 10-19, City of Rochester/Monroe County. Retrieved from Metro Council for Teen Potential website: https://metrocouncilrochester.files.wordpress.com/2012/11/final-teen-birth-rates-1990-to-2011.pdf

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.

Schochet, P. Z., & Chiang, H. S. (2011). Estimation and identification of the complier average causal effect parameter in education RCTs. *Journal of Educational and Behavioral Statistics, 36*(3), 307-345.

Ten Have, T. R., Normand, S. L., Marcus, S. M., Brown, C. H., Lavori, P., & Duan, N. (2008). Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatric Annals, 38*(12), 772-783.

Weisman, S. A., & Gottfredson, D. C. (2001). Attrition from after school programs: Characteristics of students who drop out. *Prevention Science, 2*(3), 201-205.

Wyman National Network. (2015). *Wyman Center 2014 annual report: Transforming today's teens, discovering tomorrow's leaders*. Retrieved from St. Louis, Missouri: Wyman Center, Inc. website: http://wymancenter.org/wp-content/uploads/2015/11/2014-Annual-Report.pdf

Appendix A: Data collection efforts

Table A.1. Outcome of site recruitment effort.

	Number of Sites	Notes
Total number of school-based and free-standing recreation centers serving target youth in Rochester, New York	27	
Did not meet eligibility criteria	9	*One (1) devoted exclusively to the geriatric population* *One (1) devoted exclusively to older youth* *Seven (7) provide summer programming only*
Not recruited in to study, placed on wait list	7	*Five (5) due to small size (both in physical resources and number served)* *Two (2) were school facilities with access to only a gym and pool*
Recruited in to study	11	
Sites replaced from cohort 1 to cohort 2	1	Reason for replacement: Scheduling conflicts and limited physical resources devoted to program in cohort 1 site
Sites replaced from cohort 2 to cohort 3	2	Reason for replacement: Site 1. Joint agreement between school principal and City project leadership as well as limited space for after school activities at this school (this was the replacement site in the previous year) Site 2. Joint agreement between Recreation Center director and City project leadership as the neighborhood school saw an influx of recreational programming for their students and this smaller recreation center was serving a larger segment of older youth

Table A.2. Timing of data collection efforts used in the impact analysis of TOP®.

Data collection effort	Cohort 1	Cohort 2	Cohort 3
Baseline survey	*9/10/2012 – 9/28/2012*	*9/9/2013 – 10/4/2013*	*9/10/2014 – 10/3/2014*
Start date of programming	*10/1/2012*	*10/7/2013*	*10/6/2014*
End date of programming	*5/20/2013*	*5/12/2014*	*5/22/2015*
Immediate post-intervention survey	*5/20/2013 – 7/31/2013*	*5/13/2014 – 7/28/2014*	*5/25/2015 – 7/6/2015*

Appendix B: Implementation evaluation data collection

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?	Each program facilitator completed TOP® specific fidelity monitoring logs (FMLs). The data collected include date of session, number of youth in attendance at session, duration of session, and frequency of sessions. Additionally, number of community service learning hours was tracked on attendance forms.	FMLs and attendance forms were completed weekly for each session.	TOP® program facilitators
Adherence: What and how much was received?	Each program facilitator completed attendance records and FMLs as noted above. These data were used to determine number and percent of sessions attended and percentage of sample that did not attend at all (no-shows).	Attendance records and FMLs were completed weekly for each session.	TOP® program facilitators
Adherence: Who delivered material to youth?	List of staff members, dates of trainings, and qualifications of staff members were maintained and updated at least annually. Facilitators providing the sessions and assistants were noted on the attendance forms. (Note: Data were limited by not consistently collecting information regarding assistants being present on the attendance forms for year 1 and part of year 2).	Data on all program facilitators was available to program administrative staff and evaluation team members. Attendance forms were collected weekly.	Project Director/TOP® program facilitators

Note: TPP = Teen Pregnancy Prevention.

Table B.1 (continued). 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
Quality: Quality of staff-participant interactions	The FML captured information regarding program quality as identified by Wyman (e.g., I listened more than I talked; I acknowledged and rewarded desirable behavior) Observers assessed comparable items on the observation form (e.g., The implementer's rapport and communication with participants was...)	FMLs were completed by facilitators on a weekly basis. A random selection of 10% of all sessions provided was observed annually by trained outside (non-program) observers.	TOP® program facilitators/Trained observers
Quality: Quality of youth engagement with program	The FML captured information regarding program quality as identified by Wyman (e.g., I elicited questions/responses from multiple members of the group) Observers assessed comparable items on the observation form (e.g., How actively did the group members participate in discussions and activities?)	FMLs were completed by facilitators on a weekly basis. A random selection of 10% of all sessions provided were observed annually by outside (non-program) trained observers.	TOP® program facilitators/Trained observers

Note: TPP = Teen Pregnancy Prevention.

Table B.1 (continued). 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
Counterfactual: Experiences of comparison condition	Program facilitators completed WR specific fidelity monitoring logs (FMLs) which included date of session, number of youth in attendance, duration of session, frequency of sessions, and any potential overlap (overlap in what?) with TOP® programming.	Attendance forms and FMLs were collected and reviewed monthly	WR program facilitators/Participating youth
		10% of WR sessions were observed annually.	
	WR facilitator completed attendance records and FMLs were used to determine number and percent of sessions attended and percentage of sample that did not attend at all (no-shows).	Pre and post intervention surveys are reviewed at time of data collection.	
	FMLs were completed for counterfactual condition by both program facilitators and observers.	Note: FMLs were completed inconsistently by facilitators in year 1 and year 2. Observations of WR sessions did not occur in year 1.	
	Items on post-intervention survey assessed type and amount of exposure to TPP programming		

Note: TPP = Teen Pregnancy Prevention.

Table B.1 (continued). 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
Context: Other TPP programming available or offered to study participants (both intervention and comparison)	Minutes and field notes from participation in community meetings. Survey items on questionnaire at each data collection point. At baseline and at post intervention each year, we asked youth whether he/she participates (d) in the TOP® and whether he/she participated in a Work Readiness Program. At baseline of each year, we asked youth whether they have participated in any of the local teen pregnancy/HIV prevention programs and whether they have received information on relationships, dating, marriage, or family life; abstinence from sex; methods of birth control; where to get birth control; STIs/STDs; how to talk with your partner about whether to have sex or use birth control; how to say "no" to sex; how babies are made.	Semi-annually and as needed. At time of survey administration.	Project Director Participating youth
Context: External events affecting implementation	Minutes and field notes from participation in agency and community meetings. Attendance records indicate contextual aspects of implementation.	Semi-annually and as needed. Weekly attendance records	Project Director
Context: Substantial unplanned adaptation(s)	FMLs completed by program facilitators and random observation of 10% of TOP® sessions by trained outside observers.	Materials were reviewed on a monthly basis by the evaluation team and the Project Director.	Project Director/ Evaluation team

Note: TPP = Teen Pregnancy Prevention.

Appendix C: Study sample

Table C.1. Cluster and youth sample sizes by intervention status.

Number of:	Time period	Total sample size	Intervention sample size	Comparison sample size	Total response rate	Intervention response rate	Comparison response rate
Number of CLUSTERS (sites)							
At beginning of study	11 sites randomized annually for 3 years	*33*	*18*	*15*			
Contributed at least one youth at baseline	*Baseline*	*33*	*18*	*15*	*100.0*	*100.0*	*100.0*
Contributed at least one youth at follow-up	*Immediate post-programming*	*33*	*18*	*15*	*100.0*	*100.0*	*100.0*
Number of YOUTH							
In recreation center programs and/or in catchment area who expressed some level of interest in study participation[a]		*1978*	*1124*	*854*			
Who consented prior to randomization		*1188*	*677*	*511*	*60.1*	*60.2*	*59.8*
Contributed a baseline survey	*Immediately pre-programming*	*1018*	*571*	*447*	*85.7*	*84.3*	*87.5*
Contributed a follow-up survey	*Immediately post-programming*	*961*	*562*	*399*	*80.9*	*83.0*	*78.1*
Contributed both a baseline and immediate post intervention survey with complete covariate and outcome data		*708*	*414*	*294*	*59.6*	*61.2*	*57.5*

[a] This is the number of youth who expressed some interest in participating in the study during the recruitment months. Total number of 11-14 year old youths attending or enrolled in each recreation center was not systematically tracked.

Appendix D: Analytic sample and non-analytic sample baseline differences

Table D.1. Key baseline measures for youth completing the Helping Youth THRIVE: Youth Development Survey who were in the analytic sample versus those not in the final analytic sample.

Baseline measure	n	Analytic Sample Mean or % (standard deviation)	n	Non-Analytic Sample Mean or % (standard deviation)	Mean/% difference	*p*-value of difference
Demographics						
Age	708	12.35 (1.12)	411	12.53 (1.14)	-.18	.015
Gender (female)	708	53.7%	416	29.6%	24.1%	< .001
Ethnicity: Hispanic	708	31.4%	284	27.1%	4.3%	.180
Race: White only	708	15.4%	308	14.3%	1.1%	.532
Race: Black/African-American only	708	64.5%	308	67.5%	-3.0%	.287
Race: Other race only	708	2.4%	308	4.2%	-1.8%	.183
Race: Black/African-American and White	708	8.6%	308	6.5%	2.1%	.253
Race: Black/African-American and other race(s)	708	5.6%	308	5.8%	-0.2%	.900
Race: Black/African-American and White and other race(s)	708	3.4%	308	1.6%	1.8%	.113
Sexual Outcome						
Ever had sexual intercourse	708	6.5%	369	8.7%	-2.2%	.179
Sexual Intentions Outcomes						
Intention to have sexual intercourse	708	3.09 (0.97)	259	3.04 (1.02)	.05	.007
Intention to use (or have partner use) condom	708	2.26 (1.21)	236	2.24 (1.25)	.02	.810
Intention to use effective means of birth control	708	2.32 (1.18)	209	2.29 (1.19)	.03	.582

Source: Helping Youth THRIVE: Youth Development Survey administered at baseline assessment.,

Appendix E: Equation for estimating baseline equivalence

The following model was used to test for treatment-control differences on the baseline value of each outcome measure for the primary and secondary reseach questions, as well as for the following baseline demographic measures: age, sex, ethnicity, and race. A multilevel model was used to account for the clustering of youth within out-of-school programs. For binary demographic and outcome variables, a logistic approach was used; for the continuous sexual intentions measures, a normal distribution approach was used, with robust standard errors to adjust for non-normality in the distributions of these variables.

Logistic

(1) Level 1: $log(p(x)/(1-p(x))) = \beta_{oj}$

(2) Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} \times (T_j - \bar{T}_J) + \sum_{m=1}^{M} \gamma_m A_{mj} + \sum_{n=1}^{N} \gamma nBnj + u_{0j}$

At level 1 (individual level):

$log(p(x)/(1-p(x)))$ *is the log odds of having the characteristic of interest (i.e., variable coded as 1 if characteristic is present; 0 otherwise; gender has females coded as 1)*

β_{0j} *mean value of the baseline measure in cluster j*

At level 2 (level of randomization):

γ_{00} *is the global mean of the baseline measure*

γ_{01} *is the coefficient of interest, which represents the estimated difference in log odds between the treatment and control groups*

$(T_j - \bar{T}_J)$ T_j *is a dummy variable equal to 1 if the recreation center was assigned to the treatment group; 0 if control group.* $(T_j - \bar{T}_J)$ *represents centered treatment dummy variable.*

u_{0j} *is the redidual error for recreation center j, which is assumed to be independently and identically distributed.*

Normal

(1) Level 1: $Y_{ij} = \beta_{0j} + \varepsilon_{ij}$

(2) Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} \times (T_j - \bar{T}_J) + \sum_{m=1}^{M} \gamma_m A_{mj} + \sum_{n=1}^{N} \gamma_n B_{nj} + u_{0j}$

At level 1 (individual level):

Y_{ij} *is the baseline demographic or behavioral measure for youth i in cluster j.*

β_{0j} *is the mean value of the baseline measure in cluster j*

ε_{ij} *is the residual error (variance) for student i in cluster j, which is assumed to be independently and identically distributed.*

At level 2 (level of randomization):

γ_{00} *is the global mean of the baseline measure*

γ_{01} *is the coefficient of interest, which represents the estimated difference in between the treatment and control groups*

$(T_j - \bar{T}_J)T_j$ *is a dummy variable equal to 1 if the recreation center was assigned to the treatment group; 0 if control group.* $(T_j - \bar{T}_J)$ *represents centered treatment dummy variable.*

A_{mj} *are centered dummy variables representing 3 CBO stratification variables*

B_{nj} *are centered dummy variables representing 2 cohort variables*

μ_{0j} *is the redidual error for recreation center j, which is assumed to be independently and identically distributed.*

Appendix F: Impact Model Equations Specification

The following model was used to test for treatment-control differences at the immediate post-intervention assessment point for each outcome measure for the primary and secondary reseach questions. A multilevel model was used to account for the clustering of youth within sites. For binary outcome variables, a logistic approach was used; for the continuous sexual intentions measures, a normal distribution approach was used, with robust standard errors to adjust for non-normality in the distributions of these variables. In all instances, covariates included the baseline value of the dependent variable, age of the youth at baseline, gender, ethnicity, and race (black and white versus other races) at the individual level and cohort status and CBO stratification at the recreation center level. Individual outcomes are modeled at level 1, while level 2 represents the level of cluster randomization.

Logistic:

(1)   Level 1: $\log(p(x)/(1-p(x))) = \beta_{oj} + \sum_{k=1}^{K} \beta_{kij} X_{kij}$

(2)   Level 2: $\beta_{oj} = \gamma_{00} + \gamma_{01} * (T_j - \bar{T}_j) + \sum_{m=1}^{M} \gamma_m A_{mj} + \sum_{n=1}^{N} \gamma_n B_{nj} + \mu_{oj}$

At level 1 (individual level):

$\log(p(x)/(1-p(x)))$ *is the log odds of having the characteristic of interest (i.e., variable coded as 1 if characteristic is present; 0 otherwise)*

β_{oj} *mean value in log odds of the outcome measure in cluster j with all predictor variables at their mean*

β_{kij} *is the estimated coefficient for the kth baseline characteristic for youth i in cluster j*

X_{kij} *is the grand mean centered kth baseline characteristic for youth i in cluster j (demographics coded 1 for female, hispanic, black, white)*

At level 2 (level of randomization):

γ_{00} *is the global mean of the outcome measure with all predictors at their mean*

γ_{01} *is the coefficient of interest, which represents the estimated difference in log odds between the treatment and control groups*

$(T_j - \bar{T}_j)$ T_j *is a dummy variable equal to 1 if the recreation center was assigned to the treatment group; 0 if control group.* $(T_j - \bar{T}_j)$ *represents grand mean centered treatment dummy variable.*

A_{mj} *are grand mean centered dummy variables representing 3 CBO variables*

B_{nj} *are grand mean centered dummy variables representing 2 cohort variables*

μ_{oj} *is the residual error for recreation center j, which is assumed to be independently and identically distributed.*

Normal:

(1) Level 1: $Y_{ij} = \beta_{oj} + \sum_{k=1}^{K} \beta_{bij} X_{kij} + \varepsilon_{ij}$

(2) Level 2: $Y_{ij} = \beta_{oj} + \beta_{1j}(INT_{ij} - \overline{INT}..) + \sum_{k=1}^{K} \beta_{kij} X_{kij} + \varepsilon_{ij}$

At level 1 (individual level):

γ_{ij} *is the outcome behavioral measure for youth i in cluster j.*

β_{oj} *is the mean value of the baseline measure in cluster j with all predictors at their mean*

β_{kij} *is the estimated coefficient for the kth baseline characteristic for youth i in cluster j*

X_{kij} *is the grand mean centered kth baseline characteristic for youth i in cluster j (demographics coded 1 for female, hispanic, black, white)*

k^{th} *is the residual level 1 error (variance) for student i in cluster j, which is assumed to be independently and identically distributed.*

At level 2 (level of randomization):

γ_{00} *is the global mean of the baseline measure with all predictors at their mean*

γ_{01} *is the coefficient of interest, which represents the estimated difference in between the treatment and control groups*

$(T_j - \overline{T}_j)$ T_j *is a dummy variable equal to 1 if the recreation center was assigned to the treatment group; 0 if control group.* $(T_j - \overline{T}_j)$ *represents grand mean centered treatment dummy variable.*

A_{mj} *are grand mean centered dummy variables representing 3 community agency stratification variables*

B_{nj} *are grand mean centered dummy variables representing 2 cohort variables*

μ_{oj} *is the redidual level 2 error for recreation center j, which is assumed to be independently and identically distributed.*

In all instances, the coefficient on the treatment variable, γ1, is the primary coefficient of interest. We test whether the estimate of this coefficient is statistically significant at the 5 percent level using a two-tailed test. If the estimated coefficient is statistically significant, we interpret this as evidence that offering TOP® affected the outcome. If the estimated coefficient is not statistically significant, we conclude that there is no evidence that offering TOP® affected the outcome.

For the secondary outcome concerned with risk moderational effects, a cross-level interaction term was entered by crossing the level 1 predictor of interest (sexual intercourse status at baseline or intention to have sexual intercourse in the next year) with level 2 experimental condition. The following models were used.

Logistic moderation:

(1) Level 1: $(T_j - \overline{T})$

(2) Level 2: $\beta_{1j} = \gamma_{10} + \gamma_{11} * (T_j - \overline{T}_.) + \mu_{1j}$

At level 1 (individual level):

$\log(p(x)/(1-p(x)))$ *is the log odds of having the characteristic of interest (i.e., variable coded as 1 if characteristic is present; 0 otherwise)*

$(INT_{ij} - \overline{INT}_.)$ *is the grand mean centered baseline variable for the baseline characteristic involved in the interaction term (baseline sexual intercourse or baseline intentions to have sex in the next year) for youth i in cluster j*

β_{kij} *is the estimated coefficient for the kth baseline characteristic for youth i in cluster j*

X_{kij} *is the centered kth baseline characteristic for youth i in cluster j (demographics coded 1 for female, hispanic, race)*

At level 2 (level of randomization):

γ_{00} *is the global mean in log odds of the dependent measure with all predictors estimated at their mean*

γ_{01} *is the main effect coefficient of interest, which represents the estimated difference in log odds between the treatment and control groups*

γ_{10} *is the main effect coefficient for the baseline characteristic involved in the interaction term (baseline sexual intercourse or baseline intentions to have sexual intercourse in the next year)*

γ_{11} *is the moderation effect coefficient of interest, which represents the estimated difference in log odds between the experimental condition by baseline characteristic of interest (baseline sexual intercourse or baseline intentions to have sexual intercourse in the next year)*

$(T_j - \bar{T})$ T_j *is a dummy variable equal to 1 if the recreation center was assigned to the treatment group; 0 if control group.* $(T_j - \bar{T})$ *represents grand mean centered treatment dummy variable.*

A_{mj} *are grand mean centered dummy variables representing 3 CBO variables*

B_{nj} *are grand mean centered dummy variables representing 2 cohort variables*

μ_{oj} *is the residual error from the* β_{0j} *random effect for recreation center j, which is assumed to be independently and identically distributed*

μ_{1j} *is the residual error distributed from the* β_{1j} *random effect for recreation center j, which is assumed to be independently and identically*

Normal moderation:

(1)   Level 1: $Y_{ij} = \beta_{oj} + \beta_{1j}(INT_{ij} - \overline{INT_{..}}) + \sum_{k=1}^{K}\beta_{kij}X_{kij} + \varepsilon_{ij}\gamma_{ij}$

(2)   Level 2: $\beta_{oj} = \gamma_{00} + \gamma_{01} * (T_j - \bar{T}_j) + \sum_{m=1}^{M}\gamma_m A_{mj} + \sum_{n=1}^{N}\gamma_n B_{nj} + \mu_{oj}$

                  $\beta_{1j} = \gamma_{10} + \gamma_{11} * (T_j - \bar{T}_.) + \mu_{1j}$

At level 1 (individual level):

γ_{ij} *is the outcome behavioral measure for youth i in cluster j.*

$(INT_{ij} - \overline{INT_{..}})$ *is the grand mean centered baseline variable for the baseline characteristic involved in the interaction term (baseline sexual intercourse or baseline intentions to have sex in the next year) for youth i in cluster j*

β_{kij} — is the estimated coefficient for the kth baseline characteristic for youth i in cluster j

X_{kij} — is the centered k^{th} baseline characteristic for youth i in cluster j (demographics coded 1 for female, hispanic, race)

ε_{ij} — is the residual level 1 error (variance) for student i in cluster j, which is assumed to be independently and identically distributed.

At level 2 (level of randomization):

γ_{00} — is the global mean of the dependent measure with all predictors estimated at their mean

γ_{01} — is the main effect coefficient of interest, which represents the estimated difference between the treatment and control groups

γ_{10} — is the main effect coefficient for the baseline characteristic involved in the interaction term (baseline sexual intercourse or baseline intentions to have sexual intercourse in the next year)

γ_{11} — is the moderation effect coefficient of interest, which represents the estimated difference between the experimental condition by baseline characteristic of interest (baseline sexual intercourse or baseline intentions to have sexual intercourse in the next year)

$(T_j - \overline{T})$ — T_j is a dummy variable equal to 1 if the recreation center was assigned to the treatment group; 0 if control group. $(T_j - \overline{T})$ represents grand mean centered treatment dummy variable.

A_{mj} — are grand mean centered dummy variables representing 3 CBO variables

B_{nj} — are grand mean centered dummy variables representing 2 cohort variables

μ_{0j} — is the residual error from the β_{0j} random effect for recreation center j, which is assumed to be independently and identically distributed

μ_{1j} — is the residual error from the β_{1j} random effect for recreation center j, which is assumed to be independently and identically distributed

Appendix G: Sensitivity analyses

Four forms of sensitivity analyses were conducted:

1. Imputation of baseline covariates to aid in statistical power
2. Propensity score analyses to help better assess effects of baseline nonequivalence
3. Analysis focused on delay of ever having had sex among sexually naïve youth at baseline, and
4. Analysis examining effects of inconsistent responses over time to the ever having had sexual intercourse question.

G.1. Imputation of Missing Baseline Covariates. As previously stated in the report, missing baseline covariates have implications for statistical power as most conventional statistical software programs implement listwise deletion of missing values. In our current sample, 176 youth are lost for analyses due to missing baseline covariates. There were 881 youth who provided post sexual behavior and intentions outcomes. Two imputation approaches were examined; dummy variable adjustment and multiple imputation. The dummy variable adjustment approach (Cohen & Cohen, 1985) was designed for missingness on predictor variables in a regression analysis. Here, for each missing predictor variable, a dummy variable is created to indicate whether or not data are missing on that predictor. All such dummy variables are included as predictors in the regression. Cases with missing data on a predictor are coded as having some constant value (usually the mean for continuous variables, 0's for dichotomous items) on that predictor and these values were adopted in the current analyses. Table G.1a presents baseline equivalence information for the intervention and control groups based on the dummy variable imputation of baseline covariates. Not surprisingly, these results are largely consistent with those found using the final analytic sample.

For the multiple imputation approach for the baseline covariates, auxiliary variables were used to improve the estimates of the missing values (Collins, Schafer, & Kam, 2001). That is, we included all variables collected at baseline with one exception. For the sexual behavior variable that was first asked as a "yes/no" variable and then asked for the number of times the event occurred, only the count variable was used in the imputation process; missing yes/no variables were imputed based on the value of the imputed count estimate (i.e., if count = 0, yes/no was imputed as no; if count > 0, yes/no was imputed as yes). Ten imputations were performed with resulting datasets combined as described by (Rubin, 1987). We included dummy cohort and CBO indicator variables to help account for the clustering of youth within site. We also employed Allison's (2002) approach to imputation with intervention studies which suggests imputing for each experimental condition separately and then pooling the data. Table G.1b presents baseline equivalence information for the intervention and control groups based on multiple imputation of baseline covariates and again, results are largely consistent with those found using the final analytic sample.

Table G.1c presents results of the two missing data imputation approaches for the main effects of condition on the primary and secondary outcomes. The pattern of findings is similar across these sensitivity analyses (size and direction of parameter estimates, size of p-values), even with the inclusion of the additional 176 youth. Similar results were noted for the

Table G.1a. Summary statistics of key baseline measures for the dummy variable imputation of baseline variables (n = 881).

Baseline measure	Intervention		Comparison		Intervention versus comparison mean difference	Intervention versus comparison *p*-value of difference
	n (missing)	Mean or % (standard deviation)	n (missing)	Mean or % (standard deviation)		
Demographics						
Age	514 (0)	12.27 (1.13)	367 (0)	12.49 (1.08)	-0.22	.152
Gender (female)	514 (0)	56.4%	367 (0)	48.0%	8.4%	.062
Ethnicity: Hispanic	514 (64)	32.5%	367 (45)	20.4%	12.1%	.003
Race: White only	514 (53)	15.2%	367 (33)	11.7%	3.5%	.651
Race: Black/African-American only	514 (53)	55.4%	367 (33)	63.2%	-7.8%	.381
Race: Black/African-American and White	514 (53)	8.2%	367 (33)	6.8%	1.4%	.687
Race: Black/African-American and other race(s)	514 (53)	5.1%	367 (33)	5.2%	-0.1%	.784
Race: Black/African-American and White and other race(s)	514 (53)	2.9%	367 (33)	2.7%	0.2%	.860
Race: Other race only (Reference)	514 (53)	2.9%	367 (33)	1.4%	1.5%	.150
Sexual Outcome						
Ever had sexual intercourse	514 (12)	4.3%	367 (7)	7.6%	-3.3%	.079
Sexual Intentions Outcomes						
Intention to have sexual intercourse	514 (37)	1.92 (0.90)	367 (26)	1.89 (0.95)	0.03	.296
Intention to use (or have partner use) condom	514 (51)	2.71 (1.11)	367 (33)	2.71 (1.19)	0.00	.925
Intention to use effective birth control	514 (61)	2.66 (1.06)	367 (42)	2.68 (1.15)	-0.02	.778

Source: Helping Youth THRIVE: Youth Development Survey administered at baseline assessment.,

Note: Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.

Table G.1b. Summary statistics of key baseline measures for the multiple imputation of baseline variables (n = 881).

Baseline measure	Intervention		Comparison		Intervention versus comparison mean difference	Interven-tion versus comparison *p*-value of difference
	n (missing)	Mean or % (standard deviation)	n (missing)	Mean or % (standard deviation)		
Demographics						
Age	514 (0)	12.27 (1.13)	367 (0)	12.49 (1.08)	-0.22	.116
Gender (female)	514 (0)	56.4%	367 (0)	48.0%	8.4%	.062
Ethnicity: Hispanic	514 (64)	38.1%	367 (45)	25.9%	12.2%	.006
Race: White only	514 (53)	19.7%	367 (33)	15.4	4.2%	.606
Race: Black/African-American only	514 (53)	56.4%	367 (33)	64.1%	-7.7%	.382
Race: Black/African-American and White	514 (53)	8.4%	367 (33)	7.1%	1.3%	.712
Race: Black/African-American and other race(s)	514 (53)	6.2%	367 (33)	6.2	0.0%	.849
Race: Black/African-American and White and other race(s)	514 (53)	3.5%	367 (33)	3.1%	0.4%	.778
Race: Other race only (Reference)	514 (53)	5.8%	367 (33)	4.1%	1.7%	.373
Sexual Outcome						
Ever had sexual intercourse	514 (12)	5.3%	367 (7)	8.5%	-3.2%	.146
Sexual Intentions Outcomes						
Intention to have sexual intercourse	514 (37)	1.92 (0.90)	367 (26)	1.89 (0.95)	0.03	.296
Intention to use (or have partner use) condom	514 (51)	2.71 (1.11)	367 (33)	2.71 (1.19)	0.00	.925
Intention to use effective birth control	514 (61)	2.66 (1.06)	367 (42)	2.68 (1.15)	-0.02	.778

Source: Helping Youth THRIVE: Youth Development Survey administered at baseline assessment.,

Note: Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.

Table G.1c. Sensitivity of impact analyses to addressing missing data through two approaches: dummy variable imputation and multiple imputation for the primary and secondary research questions (n = 881; 514 TOP®, 367 WR)

	Benchmark Analyses			Dummy Variable Imputation			Multiple Imputation		
		Adjusted Prevalence/Means			Adjusted Prevalence/Means			Adjusted Prevalence/Means	
	Experimental condition (TOP® = 1) parameter (p-value)	TOP®	WR	Experimental condition (TOP® = 1) parameter (p-value)	TOP®	WR	Experimental condition (TOP® = 1) parameter (p-value)	TOP®	WR
Primary Outcome									
Ever having had sexual intercourse	-.169 (.610)	.066	.073	-.341 (.231)	.081	.097	-.225 (.390)	.088	.100
Secondary Outcomes									
Intention to have sexual intercourse	.039 (.587)	1.63	1.59	-.025 (.687)	1.64	1.66	-.014 (.817)	1.64	1.65
Intention to use (or have partner use) condom	.032 (.744)	3.17	3.13	-.061 (.385)	3.19	3.13	.066 (.349)	3.19	3.13
Intention to use effective birth control	.020 (.833)	3.07	3.09	.006 (.942)	3.08	3.07	-.022 (.816)	3.07	3.09

Source: Helping Youth THRIVE: Youth Development Survey administered at immediate post-intervention.

Notes: Unstandardized parameter estimate presented. Analyses statistically control for sex, age, ethnicity/race, and baseline value of the dependent variable at the youth level and cohort and CBO stratification variables at the site level. See Table 2 for a more detailed description of each measure and section III.6.1. for a description of the impact estimation methods. Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.

moderation analyses (i.e., baseline sex missing indicator was a significant predictor of all outcomes examined), which also confirmed the benchmark findings, including the significant interaction on post intervention assessment intentions to have sexual intercourse in the next year, which was confirmed by both the dummy variable imputation and multiple imputation approaches.

G.2. Propensity Score Analyses. To help better assess the sensitivity of the findings in the presence of the baseline inequality in ethnicity ($p = .007$), propensity score analyses were conducted with the original analytic sample. Propensity score methods attempt to account for selection bias by matching on the propensity score, defined as the probability of exposure to the treatment conditional on a subject's observed baseline characteristics (Rosenbaum & Rubin, 1983). In summary, baseline variables are used to predict treatment status with the resulting equation being used to estimate a probability (i.e., the propensity score) of receiving the treatment for each participant. Matched sets of treated and untreated participants with similar values of the propensity score are then formed. The effect of treatment on outcomes is then estimated in these matched samples.

We used traditional statistical significance as well as the measure of standardized bias (SB) to assess matching quality. Standardized bias is both conceptually and computationally similar to effect size estimates. The standardized bias for the continuous and count covariates was calculated by dividing the difference in means of the covariate between the TOP® group and the control group by the pooled standard deviation. Standardized bias for the binary covariates was calculated as the differences in proportions divided by the pooled standard deviation (Harder, Stuart, & Anthony, 2010). Although clearly a rule of thumb as opposed to a strict cutoff, we aimed for a standardized bias of less than or equal to .10 to consider a covariate as balanced (Austin, 2010; Harder et al., 2010).

In the current analyses, a number of different propensity score samples were examined. In each sample, we followed the approach of Brookhart and colleagues (2006) and included the demographic variables of gender, age, race, and ethnicity, the baseline values of the dependent variables, and all other baseline variables thought to be related to the outcomes using the average value from the 10 imputed datasets, specifically mother's education, frequency of sexual intercourse in preceding three months, perceptions of the youth's neighborhood's physical environment, delinquency, individual strengths, empathy, self-esteem, negative affect, future expectations, amount of prosocial activities, school connectedness, teacher support, friend support, parental support, parental monitoring, family conflict, and peer risk. Cohort and CBO indicator variables were also included to help account for the clustering of youth within site. To help improve balance, higher order interaction terms were also entered into the propensity score equation. Nine samples were created; four based on the interaction of each variable with gender, age of youth, ethnicity of youth, and sexual history of youth; one sample with all of the interaction and higher order age term entered; and four sample removing each one of the four baseline interactions (i.e., interactions with three of the four variables entered). Similar to the multiple imputation approach above, dummy variables indicating recruitment recreation center were also entered to help account for the clustering of youth within recreation centers. One-to-many matched samples were created using caliper matching (caliper width = .20 of the standard deviation of the logit of the propensity score; Austin (2010). Among the nine samples assessed, the resulting sample using the interaction of all four variables assessed with each of the entered

into the equation produced the most substantial reduction in standardized bias from the original analytic sample. While Hispanic ethnicity and age result in SB's greater than the .10 cutoff, none of the sample had all baseline variables with SB's below .10 (particularly for these two variables) the final propensity score sample had the lowest average SB (.061) and all other samples had three or more variables with SB's greater than .10. Table G.2a presents baseline equivalence results for both the analytic sample and the final propensity score based sample (n = 514; 385 TOP®, 129 WR). In general, bias appears reduced in the propensity score sample, except for age. Table G.2b presents main effect results for the propensity score sample. Here, results align with the benchmark sample and there were no significant differences across the groups for any of the outcomes.

For the moderational analyses, since there was no variation in the baseline variable of ever having had sex for the propensity score sample (all youth reported being naïve in the final propensity score sample), no analyses concerning the interaction of ever having had sex at baseline by experimental condition could be examined. For the moderational effect of intentions to have sexual intercourse in the next year by experimental condition, there were no significant moderational effects for the four outcomes examined (ever having had sex and the three sexual intentions variables). That is, the significant moderation effect noted with the analytic sample on post intervention intentions to have sexual intercourse in the next year was no longer significant (p = .217). Figure G.2a. graphically presents these results. As visually noted, the differences between the full analytic sample and the propensity score sample appear to lie in the slope of the WR group. Specifically in the propensity score sample, WR youth who reported having definite intentions to have sexual intercourse in the next year at baseline were much less likely to have those intentions at post intervention while the full WR analytic sample were still much more likely to report the intention to have sexual intercourse in the next year at the post intervention assessment (correlation between baseline and post intervention intention to have sexual intercourse for the full analytic WR sample was .430 [.159 for TOP®]; .240 for the propensity score sample [.114 for TOP®]).

G.3. Delay of Sexual Onset Among Sexually Naïve Youth at Baseline. To further understand the program effects on initiation of sexual activity, we examined TOP® effects among sexually naïve youth at baseline. Essentially, this addresses the question of delay of sexual onset. Table G. 3a presents baseline equivalence information for the sample of youth reporting never having had sex. Again, significant imbalance remained for ethnicity. Among youth never having had sexual intercourse, there was no evidence of TOP® having an effect on the delay of sexual onset (b = -.144; p = .660) at the post program assessment, after accounting for the demographic variables of gender, age, race, and ethnicity. Here, 8.6% (n = 23/269) of the sexually naïve youth in the WR group became sexually active during the year while 6.9% (n = 27/393) of the sexually naïve TOP® youth did. Table G3b presents the findings among sexually naïve youth as well as findings regarding inconsistent responses (see below).

G.4. Effects of Inconsistent Responses Over Time to the Primary Outcome of Ever Having Had Sexual Intercourse. Two types of inconsistent responses were encountered: inconsistent responses within survey and inconsistent responses in ever having had sexual intercourse across baseline and post-intervention assessment points. Inconsistent responses within survey occurred when a youth answered that he/she had ever had sexual intercourse on Part A of the survey, but would respond that he/she did not ever have sexual intercourse to the same item in Part B or

Table G.2a. Mean difference and standardized bias statistics of key baseline measures for analytic sample (n = 708; 414 TOP®, 294 WR) and the final propensity score sample (n = 514 . 385 TOP®, 129 WR).

Baseline measure	Benchmark Analytic Sample Mean Difference	Benchmark Experimental versus Control p-values	Benchmark SB	Propensity Score Matching Mean Difference	Propensity Score Experimental versus Control p-values	Propensity Score SB
Demographics						
Age	-.220	.222	-.196	-.239	.171	-.217
Gender (female)	.084	.070	.173	.031	.695	.064
Ethnicity: Hispanic	.124	.007	.268	.074	.227	.157
Race: White only	.031	.851	.085	.009	.758	.023
Race: Black/African-American only	-.071	.494	-.149	-.012	.821	-.025
Race: Black/African-American and White	.014	.757	.048	-.018	.493	-.058
Race: Black/African-American and other race(s) (not White)	.004	.964	.016	.011	.608	.050
Race: Black/African-American and White and other race(s)	.000	.985	-.001	.003	.844	.016
Race: Other race only (Reference)	.024	.079	.154	.008	.657	.055
Sexual Outcome						
Ever had sexual intercourse	-.035	.095	-.140	.000	--	.000
Sexual Intentions Outcomes						
Intention to have sexual intercourse	.051	.586	.053	-.050	.748	-.052
Intention to use (or have partner use) condom	.009	.894	.008	.050	.890	.043
Intention to use effective birth control	-.010	.965	-.008	.034	.933	.030

Source: Helping Youth THRIVE: Youth Development Survey administered at baseline assessment.,

Notes: SB = Standardized Bias. P-values are from baseline equivalence tests, statistically controlling for cohort and CBO stratification variables at the site level. There was no variability in ever having had sex at baseline in the propensity score sample as all youth reported being sexually inactive. See Table III.3 for a more detailed description of each measure and section III.6.1. for a description of the impact estimation methods. Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.

Table G.2b. Sensitivity of impact analyses using propensity score matching to address the primary research questions.

	Benchmark Analyses			Propensity Score Analyses		
		Adjusted Prevalence/Means			Adjusted Prevalence/Means	
	Experimental condition (TOP® = 1) parameter (p-value)	TOP®	WR	Experimental condition (TOP® = 1) parameter (p-value)	TOP®	WR
Primary Outcome						
Ever having had sexual intercourse	-.169 (.610)[a]	.066	.073	-.043 (.916)	.042	.043
Secondary Outcomes						
Intention to have sexual intercourse	.039 (.587)	1.63	1.59	.048 (.607)	1.54	1.49
Intention to use (or have partner use) condom	.032 (.744)	3.17	3.13	-.038 (.695)	3.15	3.19
Intention to use effective birth control	.020 (.833)	3.07	3.09	-.124 (.292)	3.06	3.18

Source: Helping Youth THRIVE: Youth Development Survey administered at immediate post-intervention.

Notes: Analyses statistically control for age, ethnicity/race, and baseline value of the secondary outcome dependent variables at the youth level and cohort and CBO stratification variables at the site level. See Table 2 for a more detailed description of each measure and section III.6.1. for a description of the impact estimation methods. Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.
[a] Logistic regression unstandardized parameter estimate (p-value). No baseline value entered as there was no variation in the propensity score sample at baseline (all youth reported being sexually naïve).

Part C (duplicated item as a check) or the reverse. Most issues of this sort were resolved with post survey imputation (if it was inconsistent at baseline and post survey was answered No, the baseline value was recorded as No). After imputing post survey values for baseline inconsistencies, there still remained 5 youth whose sexual status was inconsistent within survey (1 youth at baseline and 4 at post). In the main analyses, the raw data from the main survey section (Part A) was analyzed. Of these 5, 4 reported being sexually active at the post assessment in Part A and 1 reported being sexually naïve (the one youth with inconsistent baseline responses reported being sexually active at post). Reversing the response was one form of sensitivity analysis. Here again, experimental condition was not related to ever having had sex at post survey assessment (b = -.111, p = .748). A second form of sensitivity excluded youth with inconsistent responses within surveys. Not surprisingly, neither of the two missing data patterns (baseline or post) changed the baseline equivalence findings (for missing baseline, Hispanic p = .007, gender p = .073, other race only p = .079, and having had sexual intercourse p = .108; for missing post, Hispanic p = .005, gender p = .068, other race only p = .080, and having had sexual intercourse p = .093). Excluding the one youth with baseline.inconsistency also did not alter substantive experimental condition findings (b = -.170, p = .605); nor did excluding youth with
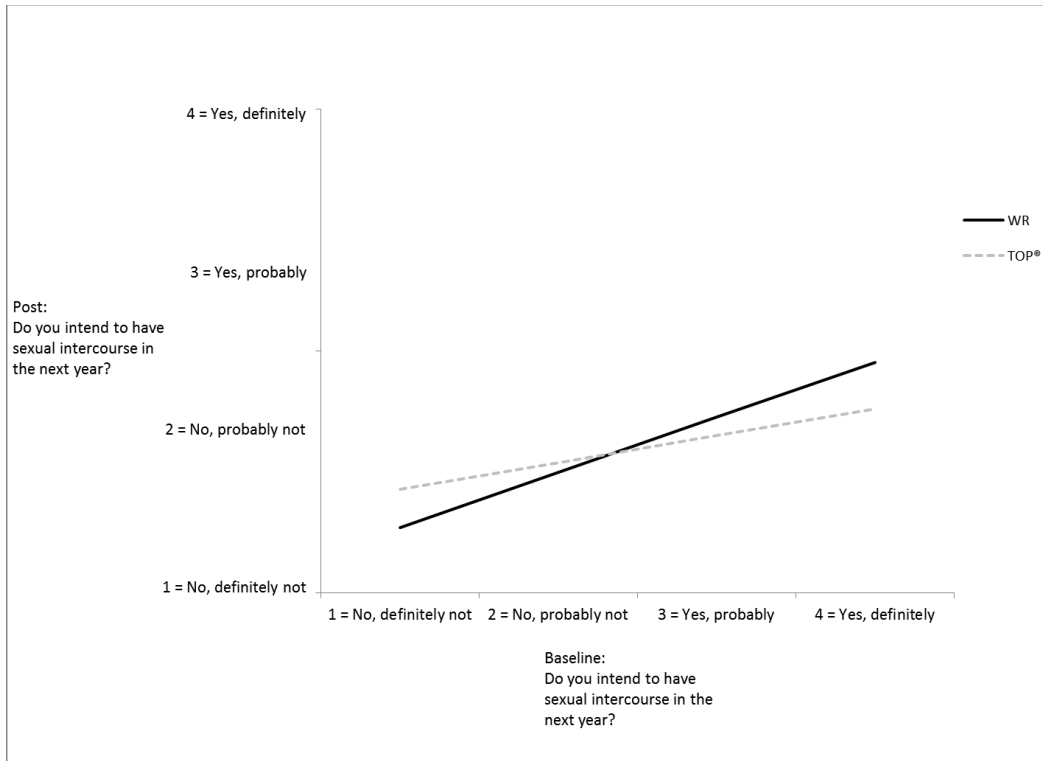
Figure G.2a. Propensity score sample baseline intentions to have sexual intercourse by condition interaction on post intentions to have sexual intercourse.

post inconsistencies (b = -.118. p = .739); nor did excluding all 5 youth (b = -.118, p = .740) . Additionally, 2 youth reported having had sexual intercourse at baseline but reported not ever having had sexual intercourse at post intervention assessment. In the analytic sample, the raw data from these 2 youth were analyzed. Treating the sexual intercourse data as missing for these two youth did not change substantive findings (b = -.144, p = .664). Baseline equivalency findings were also consistent with the main analytic sample findings (not significantly different for age, white race only, black race only, black and other race, black and white race, black and white and other race; Hispanic p = .007, gender p = .072, other race only p = .079, and having had sexual intercourse p = .106).

Table G.3a. Summary statistics of key baseline measures for youth who reported never having had sexual intercourse and who completed the immediate post intervention survey.

Baseline measure	n	Intervention Mean or % (standard deviation)	n	Comparison Mean or % (standard deviation)	Intervention versus comparison mean difference	Intervention versus comparison *p*-value of difference
Demographics						
Age	393	12.21 (1.10)	269	12.43 (1.08)	-0.22	.268
Gender (female)	393	58.3%	269	52.4%	5.9%	.261
Ethnicity: Hispanic	393	36.6%	269	25.3%	11.3%	.026
Race: White only	393	17.6%	269	14.5%	3.1%	.902
Race: Black/African-American only	393	61.0%	269	67.7%	-6.7%	.659
Race: Black/African-American and White	393	9.7%	269	7.4%	2.3%	.477
Race: Black/African-American and other race(s)	393	4.8%	269	5.6%	-0.8%	.511
Race: Black/African-American and White and other race(s)	393	3.3%	269	3.7%	0.0%	.748
Race: Other race only (Reference)	393	3.6%	269	1.1%	2.5%	.095
Sexual Intentions Outcomes						
Intention to have sexual intercourse	393	1.91 (0.95)	269	1.77 (0.93)	0.14	.450
Intention to use (or have partner use) condom	393	2.72 (1.17)	269	2.67 (1.27)	0.05	.607
Intention to use effective birth control	393	2.66 (1.14)	269	2.65 (1.24)	0.01	.722

Source: Helping Youth THRIVE: Youth Development Survey administered at baseline assessment.,

Note: Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.

Table G.3b. Sensitivity of impact analyses to delaying initiation of sexual intercourse among sexually naive youth as well as to inconsistent responses to "ever having had sexual intercourse."

	Experimental condition (TOP® = 1) parameter (p-value)	Sample Size		Adjusted Prevalence	
		TOP®	WR	TOP®	WR
Primary Outcome: Ever having had sexual intercourse					
Benchmark analysis	-.169 (.610)	414	294	.066	.073
Sexually naïve at baseline	-.144 (.660)	393	269	.047	.051
Reversing inconsistent responses within survey	-.111 (.748)	414	294	.066	.070
Baseline inconsistency treated as missing data	-.170 (.605)	414	293	.069	.075
Post inconsistency treated as missing	-.118 (.739)	412	292	.046	.050
Inconsistency within survey treated as missing	-.118 (.740)	412	291	.046	.050
Inconsistency across baseline and post treated as missing	-.144 (.664)	413	293	.123	.133

Source: Helping Youth THRIVE: Youth Development Survey administered at immediate post-intervention.

Notes: Analyses statistically control for age, ethnicity/race, and baseline value of the secondary outcome dependent variables at the youth level and cohort and CBO stratification variables at the site level. See Table 2 for a more detailed description of each measure and section III.6.1. for a description of the impact estimation methods. Sexual intention outcomes coded as 1 = No, definitely not; 2 = No, probably not; 3 = Yes, probably; 4 = Yes, definitely.

Appendix H: Implementation evaluation methods

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

Implementation element	Methods used to address each implementation element
Adherence: How often were sessions offered? How many were offered?	The number of sessions delivered was the sum of the sessions entered into SurveyMonkey. Sessions were broken down by clubs and include community service learning sessions and hours by club. Average session duration was calculated as the average (mean, median) of all provided sessions in minutes. Sessions were broken down by clubs and included average service learning hours by club. Average session frequency was calculated as the total number of sessions divided by the total number of weeks when programming was offered and was broken down by club and included service learning sessions per week by hour in addition to program sessions.
Adherence: What and how much was received?	Average numbers of sessions attended was calculated as the average number of sessions attended by youth and was categorized by club. Average number of service learning hours was the average number of hours that each youth completed and was categorized by club. Percentage of sessions attended was calculated as the total number of sessions attended by youth and divided by the total number of sessions offered to youth and categorized by club. Percentage of service learning hours was calculated as the total number of hours performed divided by the number of hours completed by youth and categorized by club. Percentage of sample that did not attend at all was calculated by total club size at beginning of programming (consented youth) divided by youth who did not attend within the first 8 weeks of programming (fidelity measure as per Wyman).
Adherence: What content was delivered to youth?	This was calculated by comparing total number of activities to be provided for the program year by each TOP® club (as per the Wyman manual) with the actual number of activities provided as reported by each TOP® club facilitator on the FMLs. A second measure assessed the total number of activities provided for the program year by each observed TOP® club (as per the Wyman manual) with the actual number of activities provided as reported by each TOP® club facilitator on the FML and confirmed by the observer on the separate FML.

Table H.1 (continued). Methods used to address implementation research questions

Implementation element	Methods used to address each implementation element
Adherence: Who delivered material to youth?	Total number of facilitators delivering the program is a simple count of facilitators and assistants implementing the program by club. We reported the average # of facilitators and assistants for all of the sessions by club. Position requirements or qualifications were calculated by comparing facilitator qualifications as noted on job application with job description for each facilitator. % of facilitators that meet or exceed job requirements was calculated as the number of facilitators who meet or exceed job requirements divided by total number of facilitators who are providing sessions. Facilitators meeting or exceeding job requirement was assessed using items in Wyman's job description, in particular the job tasks portion. % of facilitators trained was calculated as the # of facilitators who were trained divided by the total # of facilitators who delivered the program. Training refers to TOP® training and updates provided by program staff.
Quality: Quality of staff-participant interactions	Average values for the following items were calculated from the FMLs and broken down by cohort (year) and club: Did the facilitator listen more than they talked; did the facilitator acknowledge and reward desirable behavior; did the facilitator elicit questions/responses from multiple members of the group; did youth participate in setting limits and rules; and did the facilitator employ Experiential Learning Cycle techniques while facilitating the lesson. Responses from facilitators was compared to responses from the observers and congruence/dissonance was measured. Average values for the following items were calculated from the observer forms and broken down by cohort (year) and club: In general, how clear were the program implementer's explanations of activities; to what extent did the participants appear to understand the material; facilitator's level of enthusiasm; facilitator's poise and confidence; facilitator's rapport and communication with participants; facilitator's ability to effectively address questions/concerns; and facilitator's ability to demonstrate a "values neutral" approach through the lesson/activity.
Quality: Quality of youth engagement with program	Average values for the following items were calculated and broken down by cohort and club: Did youth participate in setting limits and rules; and how actively did the group members participate in discussions and activities?

Table H.1 (continued). Methods used to address implementation research questions

Implementation element	Methods used to address each implementation element
Counterfactual: Experiences of counterfactual condition	The number of sessions delivered was a sum of the sessions entered into SurveyMonkey. Sessions were broken down by clubs. Average session frequency was calculated as the total number of sessions divided by the total number of weeks when programming was offered and will be broken down by club and include service learning sessions per week by hour in addition to program sessions. Experiences of the counterfactual group included the number and % of youth who received programming that potentially could overlap with TOP® programming as noted on the FMLs completed by facilitators and observers. Responses to survey items regarding experiences of the counterfactual cohort was compared between youth who attended the counterfactual group and youth who were assigned to the control arm but did not attend the counterfactual group as well as by dosage of attendance in counterfactual group. Responses will be compared between pre- and post-survey for each group and presented as aggregate measures of central tendency and variability.
Context: Other TPP programming available or offered to study participants (both intervention and counterfactual)	All of the TPP programming available to both intervention and comparison groups available in the City of Rochester was listed in the final report. In terms of sensitivity analyses, we created subgroups of youth (e.g., control youth who report participating in a TOP® at post, youth who reported participating in some form of teen pregnancy prevention programming within the preceding 6 months of baseline) and examine outcomes within these subgroups.
Context: External events affecting implementation	All of the external events that had the potential or directly impacted implementation of TOP® was listed in the final report and was categorized by recreation center.
Context: Substantial unplanned adaptation(s)	All changes to the delivery of the TOP® curriculum noted on FMLs or by observers will be addressed in the final report.

Note: TPP = Teen Pregnancy Prevention.

Appendix I: TOP® attendance by site by cohort.

Table I.1. TOP® attendance by site by year.

	Number of Clubs	Number of Youth Enrolled	Mean Attendance % (SD)	Mean CSL Hours (SD)	Number of Zero Attendance	Attended 25 or More Sessions	Attended 19 (75%) or More Sessions	Greater than 20 CSL Hours	25 or More Sessions & 20 or More CSL Hours	19 or More Sessions & 20 or More CSL Hours
Year 1										
Site A	3	55	53.90 (43.36)	12.96 (13.51)	15	25	29	27	24	26
Site B	2	50	40.07 (39.58)	10.55 (10.50)	19	11	19	16	10	15
Site C	2	47	49.07 (39.57)	6.50 (6.56)	9	12	17	1	0	0
Site D	2	59	45.31 (37.42)	8.19 (8.70)	13	7	22	8	2	8
Site E	2	46	49.76 (36.14)	12.01 (9.36)	11	6	18	12	3	8
Site F	1	21	79.15 (27.40)	18.05 (8.77)	1	12	17	12	10	12
Total	*12*	*278*	*49.99 (39.41)*	*10.65 (10.37)*	*68*	*73*	*122*	*76*	*49*	*69*
Year 2										
Site A	3	48	51.70 (35.23)	12.49 (11.03)	6	2	16	24	2	15
Site D	2	46	50.98 (38.14)	9.60 (15.08)	8	6	20	12	6	12
Site E	2	38	53.48 (31.14)	9.00 (8.71)	3	10	14	5	5	5
Site F	1	21	72.19 (29.60)	14.62 (7.82)	1	2	14	7	2	7
Site G	1	17	70.99 (35.94)	26.35 (12.93)	0	10	11	11	10	11
Site H	1	21	73.81 (30.37)	17.33 (7.83)	1	7	13	11	6	9
Total	*10*	*191*	*58.28 (35.06)*	*13.10 (12.25)*	*19*	*37*	*88*	*70*	*31*	*59*

Note CSL = community service learning.

Table I.1 (continued). TOP® attendance by site by year.

	Number of Clubs	Number of Youth Enrolled	Mean Attendance % (SD)	Mean CSL Hours (SD)	Number of Zero Attendance	Attended 25 or More Sessions	Attended 19 (75%) or More Sessions	Greater than 20 CSL Hours	25 or More Sessions & 20 or More CSL Hours	19 or More Sessions & 20 or More CSL Hours
Year 3										
Site A	2	41	51.28 (37.92)	11.56 (10.74)	9	13	20	16	13	16
Site C	3	43	77.76 (28.34)	16.70 (7.05)	0	29	32	31	29	31
Site D	3	61	57.67 (36.05)	10.78 (8.93)	8	10	26	12	4	11
Site E	1	22	64.69 (27.03)	16.75 (7.82)	0	3	10	8	2	8
Site F	1	24	72.36 (30.43)	24.83 (11.97)	1	13	17	18	11	15
Site I	1	17	65.11 (34.93)	18.82 (7.33)	1	8	10	11	8	10
Total	*11*	*208*	*63.64 (34.34)*	*15.09 (10.14)*	*19*	*76*	*115*	*96*	*67*	*91*
Grand Total	***33***	***677***	***56.51 (37.11)***	***12.70 (11.01)***	***106***	***186***	***325***	***242***	***147***	***219***

Note: CSL = community service learning.