

Scalability of an Evidence-Based Adolescent Pregnancy Prevention Program: New Evidence From 5 Cluster-Randomized Evaluations of the Teen Outreach Program

Kimberly Francis, PhD, Susan Philliber, PhD, Eric R. Walsh-Buhi, PhD, Ashley Philliber, MS, Roopa Seshadri, PhD, and Ellen Daley, PhD

Objectives. To determine if the Teen Outreach Program (TOP), a youth development and service learning program, can reduce sexual risk-taking behaviors compared with a business as usual or benign counterfactual.

Methods. We synthesized results of 5 independent studies conducted in 5 geographically and ethnically diverse locations between 2011 and 2015 with 17 194 middle and high school students. Each study cluster-randomized classes, teachers, or schools to treatment or control groups and included the students enrolled in those clusters at baseline in an intent-to-treat analysis. Multilevel models tested impacts on recent sexual activity, recent unprotected sexual activity, and sexual initiation among the sexually inexperienced at baseline at approximately 1 and 2 years after baseline.

Results. Precision-weighted average effect sizes showed nonsignificant reductions of 1 percentage point or less in recent sexual activity (5 studies: -0.6 ; $P = .32$), recent unprotected sex (5 studies: -0.2 ; $P = .76$), and sexual initiation (4 studies: -1.1 ; $P = .10$) after 1 year.

Conclusions. There was little evidence of the effectiveness of TOP in reducing sexual risk-taking behaviors. Results underscored the importance of continually evaluating evidence-based programs that have previously been shown to be effective. (*Am J Public Health*. 2016;106:S32–S38. doi:10.2105/AJPH.2016.303307)

 See editorials, p. S5–S31.

Unplanned pregnancies and births to adolescents in the United States have been major public health issues for decades because of the potential negative consequences for both young parents and their children.¹ Although the nation's rates remain above those of most other developed countries, adolescent birth rates have declined dramatically in the United States over the past 25 years. Between 1991, when these rates were the highest in recent history, and 2014, birth rates among girls aged 15 to 19 years decreased 61%.² These declines have been observed in all states and among all racial and ethnic groups.

During this period of declining adolescent birth rates, intensive research efforts developed and tested the effectiveness of programs to prevent early pregnancies and births.³ The Teen Pregnancy Prevention

(TPP) Evidence Review systematically identifies programs that have been evaluated with at least 1 high quality research design and have shown a positive, statistically significant impact on at least 1 behavioral outcome related to adolescent pregnancy.⁴ One identified program, the Teen Outreach Program (TOP), is one of the most widely replicated TPP programs in the United States. It was offered by 68 organizations in 190 cities,

across 35 states, to an estimated 35 000 adolescents in 2014.⁵ The evidence is based on 1 randomized controlled trial in 1997 that showed a 42% lower risk of school suspension, a 39% lower risk of course failure, and a 41% lower risk of pregnancy among 342 treatment students compared with 353 control students across 25 high schools nationwide.⁶ There are no further published evaluations of the program that meet TPP Evidence Review standards.

In 2010, the Office of Adolescent Health funded 17 independent replications of TOP as part of the TPP program.⁷ Seven of these replications used randomized controlled trials to learn whether the program had impacts on pregnancy-related outcomes in a variety of settings and with populations different from those in the original study. Five of the 7 were school-based replications, and 2 were community-based.

We synthesized results from the 5 independently conducted school-based evaluations to determine if TOP affected 3 antecedents of adolescent pregnancy: onset of sexual intercourse, having recent sex, and having recent sex without using any effective method of contraception. Because the samples, settings, outcomes, and point in time of these 5 studies differed from the original study, we did not expect an exact replication of those results. Rather, the combined results of the present studies allowed us to assess the generalizability of TOP to different

ABOUT THE AUTHORS

Kimberly Francis is with the Social and Economic Policy Division at Abt Associates, Cambridge, MA. Susan Philliber and Ashley Philliber are with Philliber Research and Evaluation, Accord, NY. Eric R. Walsh-Buhi is with the Graduate School of Public Health, San Diego State University, San Diego, CA. Roopa Seshadri is with Chapin Hall at the University of Chicago, Chicago, IL. Ellen Daley is with the University of South Florida College of Public Health, Tampa.

Correspondence should be sent to Kimberly Francis, Abt Associates, 55 Wheeler Street, Cambridge, MA 02138 (e-mail: kimberly_franis@abtassoc.com). Reprints can be ordered at <http://www.ajph.org> by clicking the "Reprints" link.

This article was accepted June 4, 2016.

doi: 10.2105/AJPH.2016.303307

contexts and outcomes correlated with adolescent pregnancy. These studies provide policymakers, practitioners, and researchers with updated evidence to consider in the context of multiple replications of the program with contemporary youths.

METHODS

The 5 independent replications represented diverse settings in terms of geographic location and scope (see Table 1 for key features of each study). Three studies (Hennepin County Human Services and Public Health Department, Northwest Coalition for Adolescent Health, and The Women's Clinic of Kansas City) included both middle and high school grades, while 2 studies (Florida Department of Health and Chicago Public Schools) were implemented in high schools only. The Office of Adolescent Health required that the program replications had to be evaluated with randomized controlled trials and to measure at least 1 standard sexual behavior outcome, but it did not require that the original study design had to be replicated.

Each study was designed separately and customized to fit the local context, settings, and population.

All 5 studies used cluster randomized designs with randomization occurring within strata. Using a random number generator, 2 studies randomized classes within schools or classes within teachers and schools (Northwest Coalition and Kansas City, respectively), another randomized schools within strata defined by the majority race/ethnicity and ninth grade class size (Chicago), and another randomized teachers within schools (Hennepin County). The fifth study (Florida) randomized schools within matched pairs based on 5 characteristics in order from the most important to least important: county, courses offered, school size, region/proximity, and block/nonblock schedule. In all 5 studies, research staff conducted random assignment and study enrollment.

The studies incorporated similar eligibility criteria for the cluster-level units and students. Schools agreed to be randomly assigned (or have their teachers or classes randomly assigned) and provide 9-month long classes with class periods of at least 45 minutes to

complete a curriculum lesson. Eligible youths were those who were enrolled in the randomly assigned study classes or schools at baseline and were able to complete the surveys in the language(s) offered. Students whose parents did not provide active consent or who opted them out of the studies, as well as students who declined assent during survey administration, were excluded from the evaluation samples. Figures A–J (available as supplements to the online version of this article at <http://www.ajph.org>) show cluster and participant flow from randomization to final follow-up. The trials ended upon conclusion of the second and final follow-up data collection, and are registered on <http://www.clinicaltrials.gov> under the following trial numbers: NCT02510209, NCT02514811, NCT02514759, NCT02519530, and NCT02548871.

Each study team conducted separate a priori statistical power analyses to estimate the sample sizes needed to achieve particular minimum detectable impacts under certain assumptions (results available as a supplement to the online version of this article at <http://www.ajph.org>). Assuming a binary outcome

TABLE 1—Key Features of the Studies: Five Cluster-Randomized Evaluations of the Teen Outreach Program, 2011–2015

Feature	Hennepin County	Northwest Coalition	Kansas City	Florida	Chicago
Geographic location and scope	One urban county in MN	ID, MT, OR, WA, AK	Kansas City, MO	Nonmetropolitan counties in FL	Chicago Public School district, Chicago, IL
Study setting	24 traditional, alternative, and public charter middle and high schools in cities with highest adolescent pregnancy rates	89 traditional and alternative middle and high schools with high dropout rates in areas with high adolescent pregnancy rates	8 traditional middle and high schools in zip codes with highest adolescent birth rates	28 traditional high schools with high rates of adolescent births and STIs	44 traditional and public charter high schools in neighborhoods with highest adolescent pregnancy rates
Unit of random assignment (clusters)	63 teachers	238 classes	98 classes	28 schools	44 schools
Randomization strata	Schools	Schools	Teachers within schools	Matched school pairs	Majority race/ethnicity; 9th grade class size
Follow-up points (from baseline)	12 and 24 mo	9 and 21 mo	9 and 21 mo	9 and 19 mo	9 mo
Counterfactual	Business as usual	Alternative program	Business as usual	Business as usual	Business as usual
Comprehensive sex education required in control settings	No	No	No	No	Yes
Community service learning required in control settings	Yes: <25%	Yes: all	Yes: all	No	Yes: all

Note. STI = sexually transmitted infection.

with a mean of 50% in the control group and 80% power (0.05, 2-tailed), Hennepin County estimated 54 teacher clusters and 1080 youths would be needed to detect an 8.5 percentage point difference, Northwest Coalition estimated 460 classes and 8686 youths would be needed to detect a 4.0 percentage point difference, and Kansas City assumed 26 classes and 1884 youths would be needed to detect a 13.0 percentage point difference. Florida assumed a mean outcome of 4% with 76% power would require 60 schools to detect an effect of 2.1 percentage points. With a mean outcome of 5.5% and 90% power, Chicago assumed a sample size of 40 schools and 4600 youths to detect a difference of 1.7 percentage points.

Intervention

TOP is a youth development and service-learning program intended to reduce adolescent pregnancy and increase school success among youths aged 12 to 17 years. The program has 3 components that can be implemented in school, after school, or in community settings: (1) weekly curriculum sessions, (2) community service learning, and (3) positive adult guidance and support. All 5 replications chose to implement the program during the school day in classes that spanned the school year; one replication (Northwest Coalition) also included a small number of after school TOP classes. The intended program dosage is at least 25 weekly sessions (40–60 minutes) and 20 hours of community service learning over 9 months. Weekly sessions focus on the adolescents' service experiences (e.g., developing self-confidence, social skills, assertiveness, and self-discipline) and on a range of issues faced by students (e.g., managing family relationships, meeting new academic and employment challenges, and handling close friendships and romantic relationships).

The curriculum contains lessons at 4 developmental levels; trained facilitators select levels to match the age and developmental stage of students. Topics across levels include lessons on values clarification, relationships, communication/assertiveness, influence, goal-setting, decision-making, and human development and sexuality. Facilitators select the lessons based on needs and interests of the youths, and are required to use them in at least 80% of the sessions. Sexual health lessons

are not required, constitute less than 15% of the total curriculum, and were not predictive of outcomes in the original study.⁶ The service learning component of the program involves group or individual community service projects that are student planned and directed, and follow the planning, action, reflection, and celebration model of service learning.

Study participants in the control conditions received the “business as usual” counterfactual in 4 studies. That is, control students participated in class or school activities normally offered by the schools in the absence of TOP. The Northwest Coalition offered an alternative program that met 4 times per year and did not include any sexuality education or service learning opportunities. Most schools in these 5 replications did not require comprehensive sex education, although many did require service learning for graduation (Table 1).

Measures

All 5 studies assessed 2 outcomes with self-reported questionnaires. The measures were single-item binary indicators with identical wording: “In the past three months, have you had sexual intercourse, even once?” and “In the past three months, have you had sexual intercourse without you or your partners using any [effective] type of birth control?” Four of the 5 studies tested an additional outcome: “Have you ever had sexual intercourse?” These outcomes were not tested in the original study, which measured pregnancy on a sample of mostly female African American youths who were 16 years old on average at baseline. Outcomes were chosen because they predict adolescent pregnancy,⁸ and because the statistical power to detect impacts on these outcomes would be greater compared with the relatively rare event of experiencing or causing a pregnancy—especially for a sample with an average age of 14 years.⁹

Each study included identical self-reported measures of 3 individual-level demographic covariates: age, gender, and race/ethnicity. Race/ethnicity was included because there are wide disparities in adolescent birth rates by racial/ethnic group.² The classifications were Hispanic, non-Hispanic American Indian or Alaska Native, non-Hispanic Asian, non-Hispanic Black or African-American, non-Hispanic Native Hawaiian or other Pacific Islander, non-Hispanic White, and other.

Study participants completed questionnaires administered by research staff unaffiliated with the program implementation. Baseline surveys were conducted in schools before the intervention. First follow-up surveys took place 9 to 12 months later via a combination of paper, internet, or phone (all 5 studies used mixed modes). Second follow-up surveys occurred 19 to 24 months after baseline using the same survey modes as the first follow-up. Program staff collected attendance data, which was used to assess the extent to which youths received the full program.

Analyses

Each study used similar analytic methods. Within an intent-to-treat framework, which included all eligible youths in the sample as randomized regardless of the level of program participation or postintervention behaviors, 4 of the 5 studies used multilevel linear probability models to determine the average impact of TOP on the outcomes relative to the control group. The Chicago study used a repeated measures mixed effects model with the school and student as random effects. All 5 studies accounted for the nonindependence of students clustered within classes, teachers, or schools, and adjusted for the same student-level demographic covariates (race/ethnicity, age, gender) and student-level baseline values of the outcomes to increase statistical precision and power to detect impacts. Additional covariates were included according to the specific design features of each study (e.g., randomization strata). Using the same analytic approach, analyses of the outcome “ever had sex” were restricted to those who were sexually inexperienced at baseline (defined as those who answered “no” to the question “have you ever had sexual intercourse?”).

The effect sizes for each study were the regression-adjusted differences between the average outcomes of students in the treatment condition and students in the control condition, measured consistently across studies in percentage point units. Using a fixed-effects approach, effect sizes of each study were then weighted by the inverse variance and averaged to produce the pooled effect size for each outcome.^{10,11} Statistical significance tests of the pooled effect sizes used the *z* distribution (0.05, 2-tailed), followed by homogeneity

tests. Computing a precision-weighted average effect size across studies results in a smaller SE and, thus, more statistical power than each study on its own.¹²

RESULTS

The final analytic samples within each study contain youths with non-missing outcome data (Figures A–J). All 5 studies had low attrition at the cluster level according to the standards set by the TPP Evidence Review.¹³ Two studies, Kansas City and Chicago, had high youth attrition at first follow-up; Kansas City and Florida had high youth attrition at second follow-up. Attrition is 1 element used by the TPP Evidence Review to assess the overall quality of the evidence. All studies assessed baseline equivalence on demographics and the 2 outcome measures tested on the full sample; all studies controlled for these variables, irrespective of non-equivalence. Within each study, the analytic samples of treatment and control students at both follow-up points were statistically equivalent at baseline with some exceptions: the Florida study found the control group from the second follow-up to be more likely to have ever had sex at baseline than would the treatment group; the Northwest Coalition's second follow-up sample was non-equivalent on the percentage of Blacks/African Americans; and the Kansas City sample was nonequivalent on grade at both follow-ups and percentage of Whites at the

second follow-up. Supplementary materials, available at <http://www.ajph.org>, present the baseline equivalence tests for each study.

The baseline demographic characteristics of the analytic samples varied across studies. There were slightly more female than male participants, and the average age was 14 years. Kansas City had the highest proportion of non-Hispanic Black/African American youths (59% treatment, 60% control), followed by Chicago (48% treatment, 47% control). The highest proportion of Hispanic youths were in the Chicago study (48% treatment, 49% control), and most of the Florida study sample was non-Hispanic White (62% treatment, 64% control).

The extent to which treatment group members received the intended amount of weekly sessions and service learning hours varied. Table 2 shows the median number of sessions and service learning hours completed by treatment group youths from each study, and the percentages of who received the intended program dosage. The median number of weekly sessions received ranged from 20 to 27, although the percentage of youths who completed the minimum 25 sessions was as low as 8% in Kansas City and as high as 67% in Hennepin County. In general, lower percentages of youths completed the required 20 hours of service learning. Among the 4 studies reporting this information at the youth level, the median hours completed ranged from 3 in Kansas City and Chicago, where almost no youths completed 20 hours, to 18 hours in Northwest

Coalition and Hennepin County, where 43% and 39% of youths completed 20 hours, respectively.

At first follow-up, the average precision-weighted effect sizes were not statistically significant for any of the 3 outcomes (Table 3). For the measure of recent sexual activity, the average effect size was a difference of less than 1.0 percentage point (–0.6 percentage point; 95% confidence interval [CI] = –1.8, 0.6). Effect sizes ranged from a 5.1 percentage point reduction for the Florida treatment group to a 1.2 percentage point increase for the Northwest Coalition treatment group. The average effect size for sex without contraception was even smaller (–0.2 percentage point; 95% CI = –1.1, 0.8), with individual study effects ranging from 0.7 percentage point in the Northwest Coalition to a favorable –3.1 percentage point difference in Hennepin County. Among the sexually inexperienced at baseline, the average treatment effect was a 1.0 percentage point decrease in the likelihood of sexual initiation (–1.1 percentage point; 95% CI = –2.3, 0.2). Effect sizes among the 4 studies that contributed estimates ranged from reductions of 3.0 percentage points in Kansas City and Florida, to a 2.3 percentage point increase in the likelihood of sexual initiation for Hennepin County.

At second follow-up, the average precision-weighted effect sizes were similarly small and nonsignificant for all outcomes (Table 4). There was a slight reduction in recent sexual activity for the treatment group

TABLE 2—Amount of the Program Received by Youths in the Treatment Groups: Five Cluster-Randomized Evaluations of the Teen Outreach Program, 2011–2015

Characteristics	Hennepin County	Northwest Coalition	Kansas City	Florida	Chicago
Weekly sessions attended by youths, median	27	24	20	24	23
Attended ≥ 25 sessions, %	67	46	8	48	28
Intervention classes offering ≥ 1 sexual health lesson from curriculum, %	Not available	34	100	100	100
Service hours completed by youths, median	18	18	3	15 ^a	2.5
Completed ≥ 20 service hours, %	39	43	1	11 ^b	0
Completed 25 sessions and 20 service hours, %	35	40	1	9 ^c	0
Sample size, no.	763	3496	566	1123	3141

Note. Dosage is based on the treatment group samples used in the impact analyses.

^aNot youth-level data; number means among 70 classes, the median number of service learning hours offered was 15 hours.

^bNot youth-level data; number means 8 out of 70 classes (11%) offered ≥ 20 service learning hours.

^cNot youth-level data; number means 6 out of 70 classes (9%) offered ≥ 25 sessions and ≥ 20 service learning hours.

(-1.4 percentage point; 95% CI = -2.80, 0.08). Sex without contraception and sexual initiation among the sexually inexperienced at baseline showed unfavorable increases of less than 1 percentage point each (0.2 percentage point; 95% CI = -0.9, 1.3; and 0.4 percentage point; 95% CI = -1.4, 2.3, respectively). Tables available as a supplement to the online version of this article at <http://www.ajph.org> include the full results of the pooled effects and homogeneity tests. Data were not available from Chicago for the second follow-up.

DISCUSSION

These 5 independent studies are among the first rigorous evaluations of TOP since the original randomized controlled trial in 1997. Together, they provide valuable information

about the robustness of evidence over time and across diverse geographic regions, levels of urbanicity, and participant characteristics. Although previous studies reported that TOP reduced adolescent pregnancy,^{6,14,15} school suspension,^{6,14-16} and course failure,^{6,14-16} only 1 study met the evidence standards for high quality.¹⁷ That study found that TOP reduced self-reported adolescent pregnancy rates for female program participants.⁶ The combined results across the present studies indicated no statistically significant advantage for the intervention over the control groups for 3 sexual risk behaviors at either follow-up. That these studies were conducted almost 20 years after the original study and in different contexts means that TOP might have been effective at that point in time in that particular study context, but its effectiveness and generalizability should be reconsidered in light of new evidence.

There were several potential reasons for the general absence of effects. First, impacts of interventions on sexual risk-taking behaviors might be more difficult to detect among today's youths because the prevalence of these behaviors has declined over time. For example, in 1995, 53% of US high school students reported ever having had sexual intercourse and 9% reported first sex before the age of 13 years.¹⁸ Almost 20 years later, however, 47% of youths reported ever having had sex, and 6% reported early sex.¹⁸ Moreover, compared with youths in 1995, a greater percentage of youths in 2013 reported using condoms and birth control pills at last sexual intercourse, and a smaller percentage of youths reported having 4 or more sex partners and sex with a person in the last 3 months.¹⁸ As the prevalence of a targeted behavior declines, the statistical power to detect impacts also declines.⁹ However, this explanation was less plausible because the studies we described focused on relatively prevalent outcomes, had relatively large sample sizes, and combined effects to increase statistical power.

Second, a strong treatment-control contrast might be more difficult to achieve now than in the 1990s. Among the present studies, the control condition in all but Florida required at least some service learning. However, only the Chicago study schools required comprehensive sex education. As such, the gap in service learning engagement between the intervention and control groups might be smaller than in the original implementation, which could have compromised the ability to detect any impact of the program.

Third, the absence of effects might be caused by the program implementation approaches. For instance, treatment group youths involved in the original TOP study emphasized individualized service placements (e.g., there was a dedicated service learning coordinator to help youths plan their service experiences), whereas youths involved in the present study might have been offered less intimate service opportunities through which they could personally meet those they served and see tangible benefits of their work. The extent to which the quality of the service experience was related to program impacts is an area in need of further study. Also, most of the treatment group youths in these studies did not receive the intended amount of

TABLE 3—Results at First Follow-up: Five Cluster-Randomized Evaluations of the Teen Outreach Program, 2011–2015

	Ever Had Sex ^a	Total	Without Contraception
Hennepin County			
T (n = 763), %	10.0	14.3	4.1
C (n = 460), %	7.7	15.3	7.2
Effect ±SE	2.3 ±0.02	-1.0 ±0.03	-3.1 ±0.03
Northwest Coalition			
T (n = 3496), %	12.5	24.7	9.7
C (n = 3279), %	11.6	23.5	9.0
Effect ±SE	0.9 ±0.01	1.2 ±0.01	0.7 ±0.01
Kansas City			
T (n = 566), %	6.9	10.5	3.5
C (n = 447), %	10.1	10.4	3.6
Effect ±SE	-3.2 ±0.02	0.1 ±0.02	-0.1 ±0.01
Florida			
T (n = 1123), %	11.7	16.1	7.0
C (n = 1427), %	15.1	21.3	9.0
Effect ±SE	-3.4 ±0.01	-5.1 ±0.01	-2.0 ±0.01
Chicago			
T (n = 3141), %	NA	21.8	3.2
C (n = 2492), %	NA	21.3	3.0
Effect ±SE	NA	0.5 ±0.03	0.2 ±0.03
Weighted average effect (95% CI)	-1.1 (-2.3, 0.2)	-0.6 (-1.9, 0.6)	-0.2 (-1.2, 0.8)

Note. C = control group; CI = confidence interval; NA = not available (outcome not tested); T = treatment group. Mean percentages, effects, and SEs are regression-adjusted. Effects are percentage point differences in adjusted means. Subgroup sample sizes were n = 973 (Hennepin County), n = 3993 (Northwest Coalition), n = 914 (Kansas City), and n = 2347 (Florida).

^aSubgroup sample of sexually inexperienced at baseline.

TABLE 4—Results at Second Follow-up: Four Cluster-Randomized Evaluations of the Teen Outreach Program, 2011–2015

Variables	Ever Had Sex ^a	Had Sex in Last 90 Days	
		Total	Without Contraception
Hennepin County			
T (n = 751), %	20.0	16.8	6.3
C (n = 445), %	14.7	19.1	6.6
Effect ±SE	5.3 ±0.02	-2.3 ±0.03	-0.3 ±0.02
Northwest Coalition			
T (n = 3114), %	22.3	26.2	8.6
C (n = 2944), %	21.5	27.1	7.5
Effect ±SE	0.8 ±0.01	-0.9 ±0.01	1.1 ±0.01
Kansas City			
T (n = 522), %	18.6	15.5	5.6
C (n = 410), %	19.2	15.4	5.1
Effect ±SE	-0.6 ±0.03	0.1 ±0.02	0.5 ±0.02
Florida			
T (n = 982), %	26.2	22.3	8.6
C (n = 1232), %	28.3	25.5	10.3
Effect ±SE	-2.1 ±0.02	-3.2 ±0.02	-1.8 ±0.01
Weighted average effect (95% CI)	0.4 (-1.4, 2.3)	-1.4 (-2.90, 0.07)	0.2 (-0.9, 1.4)

Note. C = control group; CI = confidence interval; T = treatment group. Mean percentages, effects, and SEs are regression-adjusted. Effects are percentage point differences in adjusted means. Second follow-up data for Chicago were not reported. Subgroup sample sizes were n = 956 (Hennepin County), n = 3792 (Northwest Coalition), n = 839 (Kansas City), and n = 1934 (Florida).

^aSubgroup sample of sexually inexperienced at baseline.

weekly sessions and service learning hours. This was especially true for the Kansas City and Chicago studies, where treatment youths received virtually no service learning. However, the explanation cannot be as simple as low dosage. The Florida study, which took place exclusively in non-metropolitan settings, found the largest effect sizes consistently favoring the treatment group although less than 10% of the classes offered the full dosage (Table 2). It might be that even small doses of service learning and sexual health education in nonmetropolitan settings with fewer community resources was enough to generate small impacts. Non-experimental research on service learning in the context of TOP suggested that the quality of the experience was more predictive of positive outcomes than was the number of hours spent on service learning; the dosage necessary to achieve impacts remains unknown.¹⁵

Lastly, it was possible that the intervention, as currently designed, was not effective, regardless of context. This raises the question of

whether a youth development strategy, valued for its flexible, youth-centered approach to providing opportunities for engagement and positive relationships with adults and peers,¹⁹ alone can affect sexual behaviors without a more direct, consistent focus on sexual health.

Limitations

This research had several limitations. First, our results might have potential self-report bias from participants. Second, the final analytic samples consisted of youths with nonmissing outcome data, and the results of each individual study were only generalizable to the participating schools and students. Third, as noted previously, although the program was mostly offered in accordance with fidelity standards, many youths did not receive the full dosage. Lastly, subgroup analyses were limited to baseline sexual experience. Despite these limitations, our samples included a large number of youths with racial/ethnic diversity and who were from both rural and urban settings.

Furthermore, each study maintained a rigorous randomized controlled trial design.

Conclusions

Based on data from 5 studies that, together, included more than 17 000 youths in 5 diverse geographic settings, we found little evidence to support the effectiveness of TOP in reducing sexual risk-taking behaviors that should, in turn, reduce adolescent pregnancy. Although these specific outcomes were not measured in the original study,⁶ our findings expand the evidence base on this popular program by measuring its impact on the sexual risk-taking behaviors that lead to pregnancy. The general lack of evidence underscores the importance of continuing to evaluate evidence-based programs that were shown to be effective at a particular point in time, in a specific implementation context, and with specific populations. Because most programs identified by the TPP Evidence Review as of 2016 are based on evidence from single studies, the extent to which these programs will be effective in different settings and with different populations over time is a critical question as the evidence base continues to evolve. *AJPH*

CONTRIBUTORS

All authors reviewed drafts and approved the final version. K. Francis conceptualized the article, led the writing, synthesized the results of the 5 studies presented in the article, and was responsible for design, data collection, and analysis of the Hennepin County study. S. Philliber contributed to the direction and writing of the article, and was the principal investigator for the Northwest Coalition study. E. R. Walsh-Buhi contributed to the direction and writing of the article, and was responsible for design, data collection, and analysis of the Florida study. A. Philliber was responsible for data collection and analysis for the Kansas City study, and data collection for the Northwest Coalition study. R. Seshadri was responsible for the design, data collection, and analysis of the Chicago study. E. Daley was responsible for the analysis of the Florida study.

ACKNOWLEDGMENTS

This publication was prepared under grant numbers TP1AH000078, TP1AH000075, TP1AH000021, TP1AH000017-01, and TP1AH000066 from the Office of Adolescent Health, US Department of Health and Human Services. This article is based on 5 larger reports located on the Office of Adolescent Health's Web site.

Note. The views expressed in this report are those of the authors and do not necessarily represent the policies of the Department of Health and Human Services or the Office of Adolescent Health.

HUMAN PARTICIPANT PROTECTION

Each of the 5 studies received institutional review board approval independently.

REFERENCES

1. Hoffman SD. Consequences of teen childbearing for mothers: updated estimates of the consequences of teen childbearing for mothers. In: Hoffman SD, Maynard RA, eds. *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy*. 2nd ed. Washington, DC: Urban Institute Press; 2008:74–92.
2. Hamilton BE, Martin JA, Osterman MJ, Curtin SC. *Births: Preliminary Data for 2014*. National Vital Statistics Reports. Hyattsville, MD: National Center for Health Statistics. 2015. Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_06.pdf. Accessed August 19, 2015.
3. Boonstra HD. What is behind the declines in teen pregnancy rates? *Guttmacher Policy Rev*. 2014;17(3):15–21.
4. Goesling B, Lugo-Gil J, Lee J, Novak T. *Updated Findings From the HHS Teen Pregnancy Prevention Evidence Review: April 2013 Through July 2014*. Princeton, NJ: Mathematica Policy Research; 2015.
5. The Wyman Center. *What Changing a Life Feels Like*. Available at: <http://www.wymancenter.org/nationalnetwork/TOP>. Accessed August 19, 2015.
6. Allen JP, Philliber S, Herrling S, Kuperminc GP. Preventing teen pregnancy and academic failure: experimental evaluation of a developmentally based approach. *Child Dev*. 1997;68(4):729–742.
7. Kappeler EM, Farb AF. Historical context for the creation of the Office of Adolescent Health and the Teen Pregnancy Prevention Program. *J Adolesc Health*. 2014; 54(3, suppl):S3–S9.
8. Kirby D. *Emerging Answers 2007—New Research Findings on Programs to Reduce Teen Pregnancy—Full Report*. Washington, DC: The National Campaign to Prevent Teen and Unplanned Pregnancy; 2007.
9. Schochet PZ. Statistical power for school-based RCTs with binary outcomes. *J Res Educ Eff*. 2013;6(3):263–294.
10. Lipsey MW, Wilson DB. *Applied Social Research Methods Series. Practical Meta-analysis*. Vol. 49. Thousand Oaks, CA: Sage Publications; 2001.
11. Dong N, Maynard RA, Perez-Johnson I. Averaging effect sizes within and across studies of interventions aimed at improving child outcomes. *Child Dev Perspect*. 2008; 2(3):187–197.
12. Cohn LD, Becker BJ. How meta-analysis increases statistical power. *Psychol Methods*. 2003;8(3):243–253.
13. What Works Clearinghouse. *Procedures and Standards Handbook, Version 3*. Washington, DC: US Department of Education; 2014.
14. Allen JP, Philliber S. *Evaluating Why and How the Teen Outreach Program Works: Years 3–5 of the Teen Outreach National Replication (1986/87–1988/89)*. New York, NY: Association of Junior Leagues Inc.; 1991.
15. Allen JP, Kuperminc G, Philliber S, Herre K. Programmatic prevention of adolescent problem behaviors: the role of autonomy, relatedness, and volunteer service in the teen outreach program. *Am J Community Psychol*. 1994;22(5):595–615.
16. Allen JP, Philliber S, Hoggson N. School-based prevention of teen-age pregnancy and school dropout: process evaluation of the national replication of the Teen Outreach Program. *Am J Community Psychol*. 1990;18(4): 505–524.
17. Office of Adolescent Health, Office of Public Health and Science, US Department of Health and Human Services. *Research Evidence for Teen Outreach Program (TOP)*. 2012. Available at: <http://tpevidencereview.aspe.hhs.gov/pdfs/TeenOutreachProgram.pdf>. Accessed August 10, 2015.
18. Centers for Disease Control and Prevention. *Trends in the Prevalence of Sexual Behaviors and HIV Testing, National YRBS: 1991–2013*. Division of Adolescent and School Health. Available at: http://www.cdc.gov/healthyouth/data/yrbs/pdf/trends/us_sexual_trend_yrbs.pdf. Accessed August 19, 2015.
19. Gavin LE, Catalano RF, Markham CM. Positive youth development as a strategy to promote adolescent sexual and reproductive health. *J Adolesc Health*. 2010; 46(3, supplement):S1–S6.