



The Effectiveness of Teen Pregnancy Prevention Programs: A Meta-Analysis

Russell Cole, Julieta Lugo-Gil, and Andrei Streke

KEY POINTS

- The Teen Pregnancy Prevention Evidence Review (TPPER) has systematically reviewed evaluation research on the effectiveness of teen pregnancy prevention programs (TPP) since its inception in 2009. TPPER summarizes the magnitude of program impacts in high- or moderate-quality studies - those deemed to have credible evidence. Studies reviewed by TPPER tend to be large, well-powered evaluations of TPP programs that are intended to address gaps in the literature.
- The TPPER conducted a meta-analysis of 99 high- or moderate-quality studies of 79 TPP programs, that included over 85,000 youth. The TPPER estimates that these programs are effective, on average, at improving sexual behavior outcomes for adolescents.
- On average, the TPP programs reviewed by TPPER have an effect size of 0.09, which can be conceptualized as a reduction in prevalence of risky sexual behavior by 4.3 percentage points, from 50 percent to 45.7 percent.
- The average effect size of TPP programs varies little by outcome domain, and the findings for sexually transmitted infections (STIs)/HIV, contraceptive use, and sexual activity domains were statistically significant and positive.
- The average impacts did not vary significantly across sample characteristics (for example, gender of participants), or study characteristics (for example, follow-up period).
- Impacts were largest in clinic-based programs.

INTRODUCTION

In 2009, the U.S. Department of Health and Human Services (HHS) launched the Teen Pregnancy Prevention Evidence Review (TPPER), a systematic review of research on programs (and program components) designed to reduce rates of teen pregnancy, sexually transmitted infections (STIs), and associated sexual risk behaviors. The main purpose of the review is to identify, assess, and rate the rigor and findings from evaluation studies of teen pregnancy prevention (TPP) programs.¹ The review's findings are used to describe the strength of evidence supporting different programs and to identify programs with evidence of effectiveness in reducing teen pregnancy, STIs, or associated risk behaviors. The TPPER currently includes 48 active programs with

¹ In 2023, TPPER updated its scope to review the research on program components. At the time of this brief, TPPER has only reviewed a small number of studies that estimate the effects of components. As a result, this brief focuses on the evidence for programs, where TPPER has reviewed far more evidence, and not components.

evidence of effectiveness for improving sexual behavior outcomes ([Forrester et al. 2024](#)), and information about each of these programs (and their evidence) is available on the [TPPER website](#).²

The TPPER website provides useful information for organizations interested in selecting (or comparing) among individual evidence-based programs. However, to date TPPER has not provided comprehensive information for researchers and other audiences about the average effects of TPP programs, across outcomes, populations, program types, or post-intervention follow-up periods.

A recent meta-analysis began to address this gap by examining the evidence from federally-funded TPP program evaluations that occurred between 2015 and 2019 (Juras et al. 2022). In their analysis of 52 studies, the authors found that the average impacts on sexual behavior outcomes were small and not significantly different from zero. Although that 2022 meta-analysis provided information on the effectiveness of a specific subset of TPP programs, TPPER has reviewed evidence from a substantially larger pool of programs and studies, including non-federally funded research, and evidence that has been published earlier and later than the research examined in Juras et al. (2022). Some audiences may find it valuable to understand the average effects of TPP programs when considering the wider body of research that TPPER has examined.

This brief provides a substantive contribution to the field by including this larger body of evidence in a comprehensive meta-analysis. The meta-analytic findings described here document the average program effects and the range in the effects of all findings currently found as eligible in TPPER. This large body of effectiveness research provides credible evidence because it is based on evaluation studies that received a high- or moderate-quality study rating, is based on data collection that has taken place since 2004, and includes several large evaluations of programs intended to address known gaps in the literature. We present information about the variation in effectiveness of TPP programs across outcomes, program types, subgroups, and at different follow-up periods.

This meta-analysis also benefits researchers because they can use information about the magnitude of program effects reported here (both the average impact estimate and the distribution around that average) to inform statistical power calculations as part of TPP program evaluation design work (see Moreno and Cole 2023 for a technical assistance brief on conducting power calculations in the TPP field). Researchers could also use this information to help with interpretation of their estimated impacts in relation to impacts estimated in other evaluations of similar programs; one way in which researchers do this is with Bayesian interpretation of findings.³

DATA SOURCES

A high-level summary of the TPPER process for identifying research, screening and reviewing studies, and documenting the evidence on credible studies is available in the Technical Appendix (Streke et al. 2024). Based on that process, TPPER has assembled a large database of information about effectiveness studies that have been reviewed since TPPER began in 2009. The current TPPER review protocol (version 7.0) uses a 20-year window to define eligible research, so this meta-analysis includes research in which the data were collected between 2004 and 2022 (the most recent year when data were collected in the included studies).⁴

² There are an additional 15 programs with evidence of effectiveness that are now considered to be inactive because the research is more than 20 years old or because the programs are no longer distributed.

³ Bayesian interpretation of findings (Deke and Finucane 2023; Deke et al. 2022) is an approach researchers use to assess whether an impact estimate is credible, by calculating the probability that an impact of the size of the estimated impact is favorable (or larger than some other threshold of interest) based on information from the distribution of the impacts estimated in other evaluations of similar programs.

⁴ This database is available for download on the [TPPER website](#).

This study includes 618 effect size findings from 99 high- or moderate-quality studies of 79 programs, with data collected in total from more than 85,000 youth (Exhibit 1).⁵ These records represent findings that TPPER deems *credible* because they received a high- or moderate-quality rating. Not all of these studies found statistically significant impacts, however. There are 48 programs with credible studies that show evidence of effectiveness and 31 additional programs that have credible studies but that did not show evidence of effectiveness (for a total of 79 programs with credible evidence).

Exhibit 1. Characteristics of studies and findings in the TPPER effect size database

| Characteristic | Number of studies | Percent of studies | Number of findings | Percent of findings |
|--|-------------------|--------------------|--------------------|---------------------|
| Study Rating | | | | |
| High Quality | 66 | 66.7% | 420 | 68.0% |
| Moderate Quality | 33 | 33.3% | 198 | 32.0% |
| Sample^a | | | | |
| Full sample | 92 | 92.9% | 473 | 76.5% |
| Subgroup ^b | 28 | 28.3% | 145 | 23.5% |
| Sample characteristics | | | | |
| Young women only | 28 | 28.3% | 190 | 30.7% |
| Young men only | 5 | 5.1% | 25 | 4.0% |
| All genders | 65 | 65.7% | 387 | 62.6% |
| Other | 1 | 1.0% | 16 | 2.6% |
| Outcome domain^a | | | | |
| Contraceptive use | 64 | 64.6% | 229 | 37.1% |
| Number of sexual partners | 22 | 22.2% | 69 | 11.2% |
| Pregnancy | 18 | 18.2% | 46 | 7.4% |
| Sexual activity | 76 | 76.8% | 244 | 39.5% |
| STIs or HIV | 13 | 13.1% | 30 | 4.9% |
| Follow-up timing of data collection^a | | | | |
| Short term (< 6 months) | 59 | 59.6% | 208 | 33.7% |
| Medium term (6 – 11.9 months) | 51 | 51.5% | 201 | 32.5% |
| Long term (12 – 17.9 months) | 37 | 37.4% | 157 | 25.4% |
| Very long term (18+ months) | 9 | 9.1% | 52 | 8.4% |
| Program Type | | | | |
| Clinic-based | 9 | 9.1% | 43 | 7.0% |
| Sexual health education | 61 | 61.6% | 402 | 65.0% |
| Positive youth development | 16 | 16.2% | 67 | 10.8% |
| Healthy relationships | 3 | 3.0% | 28 | 4.5% |
| Sexual risk avoidance | 10 | 10.1% | 78 | 12.6% |
| Data collection timing | | | | |
| Since 2009 | 83 | 83.8% | 504 | 81.6% |
| Between 2004 and 2009 | 16 | 16.2% | 114 | 18.4% |
| Total | 99 | | 618 | |

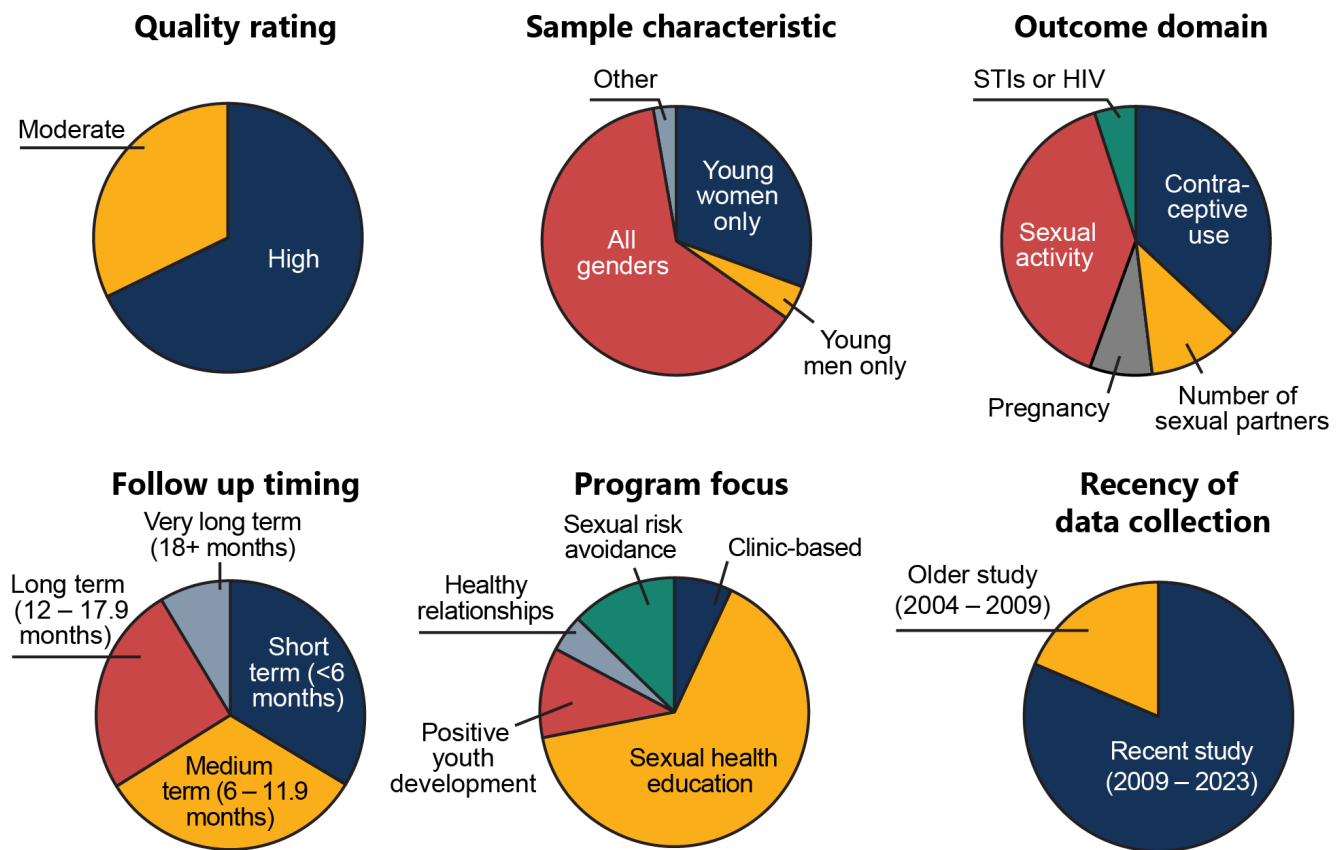
^a The number of studies for this category sums to greater than 99 (or greater than 100%) because several studies report findings in multiple categories.

^b TPPER-eligible subgroups are defined by either gender or sexual experience at baseline.

⁵ Note that the analysis eliminated findings from 10 studies where an effect size or standard error of that effect size could not be calculated; see the Technical Appendix for details on these eliminated findings.

Each record in the database includes the following information: (1) name of the program being examined along with some details about it, (2) study characteristics (for example, study design, sample characteristics), (3) contrast examined (for example, full sample or subgroup, time period when the contrast was observed), (4) outcome measure description (and associated outcome domain), (5) a standardized effect size estimate and standard error calculated from impact estimates and sample size information reported in the relevant study. (See Technical Appendix for details on how the effect size estimates and standard errors were calculated.) See Exhibit 2 for a brief summary of the characteristics of the studies included in the analysis.

Exhibit 2. Findings included in the meta-analysis



Note: Findings rated low quality are not deemed credible, and therefore not reported in TPPER nor included in this meta-analysis.

STUDY METHODS

The meta-analytic approach used for this study pools evidence across multiple outcomes, studies, and programs. As a result, we provide a summary of the effectiveness of TPP programs across outcomes and studies generally, and do not discuss the effectiveness of individual programs. (See the [TPPER website](#) for evidence about individual programs.)

We used meta-analytic approaches to calculate average effect size statistics.⁶ Meta-analysis can be conceptualized as an averaging of effect sizes, where studies with more precise impact estimates (often from larger samples) receive greater weight in the calculation of the overall average. In addition to producing an average effect, meta-analytic approaches also provide a standard error of that average effect, which is used to

⁶ TPPER calculated standardized effect sizes as Hedges' *g* when feasible, or a Cox Index for dichotomous outcomes. See the Technical Appendix for details.

create a 95 percent confidence interval (a range of values that likely contains the true average impact). Meta-analytic approaches also include an inferential test to show whether the estimated effect is statistically different from zero. The Technical Appendix provides details about the random effects analysis with robust variance estimates, which we used to pool evidence when calculating averages (Streke et al. 2024).

In this brief, we first report an overall average effect, calculated by pooling all 618 effect size findings from the TPPER database as of August 2024. This overall average effect spans all five TPPER-eligible outcome domains, all sample types, all follow-up periods, all program types, etc. This approach provides an overall estimate of the average effect of TPP programs that can serve as an easily understood and interpreted benchmark. Note: as detailed in the Technical Appendix, all effect size estimates were scaled so that a positive number represents a favorable program impact (for example, a positive effect size would represent *reduced* rates of risky sex, or a *lower* number of sexual partners).

We then report findings for each of the following five outcome domains: (1) contraceptive use, (2) STIs or HIV, (3) sexual activity, (4) pregnancy, and (5) number of partners. This approach limits the findings included in the estimate to a variety of outcome measures relevant exclusively to a particular outcome domain—for example, we combine impact findings for sexual initiation and recent sex when reporting average effects for the broader sexual activity domain. Because each of these estimated average program effects is specific to one outcome domain, they provide more detailed information than the overall program average effect described earlier. However, these domain-specific estimated average program effects pool across multiple sample types (that is, they pool across findings for full samples and subgroups) and follow-up periods (that is, they pool across findings for follow-up data collection that occurred anytime at or after program end, such as 6, 18, and 24 months after programming) which sacrifices some level of detail. (More detailed estimates of average effects are in the Technical Appendix).

We further differentiate the average program effects by reporting impacts by characteristics of the sample, characteristics of the evaluation study, and characteristics of the program.

- Characteristics of the sample we examined include (1) whether the impact was estimated using the full sample or a sample for an eligible subgroup (eligible subgroups are those defined by either gender or sexual initiation at baseline), and (2) the gender makeup of the sample included in the analysis (for example, young women only or youth of all genders).
- The characteristic of the evaluation study we examined was the data collection follow-up period (such as short-term and longer-term follow-up).
- Finally, we examined average program effects by the program type, for example, sexual health education programs and positive youth development programs.

We first conducted inferential tests to see whether impacts varied significantly across these factors. We do not present combinations of each small cross-cut of these factors within the TPPER effect size data. For example, we do not present findings specific to full-sample, short-term impacts for sexual health education programs given small sample sizes in these individual cells and non-significant differences observed in most of these inferential tests. When we *were* able to estimate average effects for combinations of sample, study, and program characteristics, these findings are available in the appendix. For all analyses, we report an average effect size statistic, a 95 percent confidence interval, and an indicator of statistical significance (based on the *p*-value)⁷ associated with the estimated average effect.

⁷ A *p*-value is the probability that an effect at least as large as the one observed would occur by chance if, in fact, there is no true effect in the population.

In the Technical Appendix, we include additional descriptive statistics for each of these analyses (such as box and whisker plots) (Streke et al. 2024). These descriptive presentations are intended to be a simple, straightforward summary of the underlying effect size data. This Technical Appendix also includes information on the distributions of the effect size data beyond average effect estimates, which could be useful for conducting power calculations (Moreno and Cole 2023) and interpreting estimated impacts in relation to the impacts estimated in other evaluations of similar programs (Deke and Finucane 2023; Deke et al. 2022).

META-ANALYSIS FINDINGS

Average overall effect size across TPP programs is positive (favorable) and statistically significant

The average effect of TPP programs assessed by TPPER was positive and statistically significantly different from zero (0.09 standard deviations, which is an effect size of 0.09).⁸ This difference can be conceptualized as a 4.3 percentage point effect relative to a 50 percent prevalence rate.⁹ For example, if the rate of risky sex in the control group is 50 percent, the average effect of a TPP program would be to reduce that rate by 4.3 percentage points to 45.7 percent.

The scientific community does not have a universal, agreed-upon threshold at which effect sizes are considered meaningful, but prior versions of the Department of Education’s [What Works Clearinghouse](#) standards and procedures (What Works Clearinghouse 2017), characterized effects of 0.25 standard deviations or larger as “substantively important.” Using this lens for TPPER, we would categorize this average effect size of 0.09 as representing a relatively small impact. For the TPP field, few findings were larger than 0.25 standard deviations—only 24 percent of observed effect sizes were this large or larger in terms of absolute value. As shown in the appendix, aside from a small number of outliers, nearly all effect size estimates ranged between -0.4 and 0.6, and the standard deviation of the observed effect sizes was 0.29.

Average effect size of TPP programs varies little by outcome domain, and findings for STIs/HIV, contraceptive use, and sexual activity domains were statistically significant and positive

The meta-analysis shows, on average, a statistically significant effect on three of the five outcome domains included in the TPPER: STIs/HIV, contraceptive use, and sexual activity. Specifically, the average effect size for interventions that examined STI outcomes was 0.20 ($p < .01$), the average effect size for interventions that examined contraceptive use was 0.08 ($p < .01$), and the average effect size for interventions that examined sexual activity outcomes was 0.07 ($p < .01$). The average impacts on pregnancy outcomes (effect size = 0.07, $p = .15$), and number of partners (effect size = 0.08, $p = .31$) were similar in size (or smaller) but neither were statistically significantly different from zero.

Exhibit 3 is a graphical representation of the average effect of TPP programs, both overall and by outcome domain, highlighting variation in program effectiveness across outcome domains. The figure presents the average effect as a dot, and a 95 percent confidence interval as lines around the dot. (Wider confidence intervals represent greater noise or uncertainty in the estimated average.) When the line does not cross the vertical effect size line at 0, the average effect is statistically significantly different from zero.

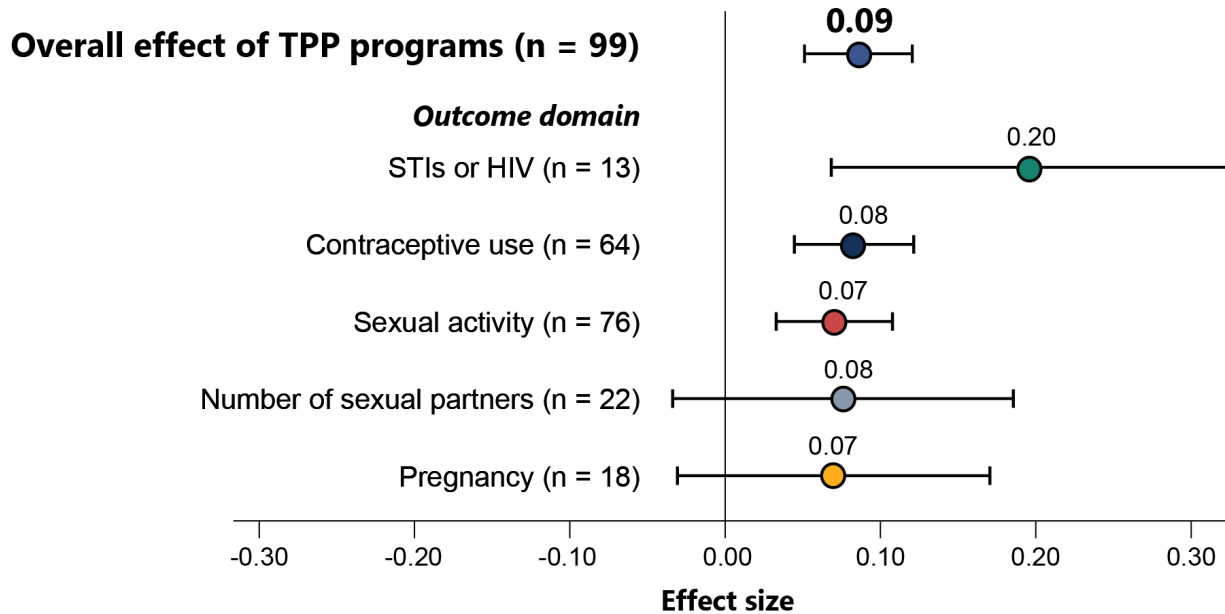
As shown in Exhibit 3, the overall (pooled across all outcomes) average effect of TPP programs is statistically significantly different from zero; the confidence interval for the overall average is also the smallest, reflecting

⁸ In this meta-analysis, an effect size with a positive sign indicates the estimated program effect is favorable to the intervention group, while an effect size with a negative sign indicates the estimated program effect is unfavorable to the intervention group.

⁹ A standardized effect size can be transformed into a percentage point difference by multiplying it against the standard deviation of a binary variable ($s = \sqrt{p(1-p)}$).

that this is the most precise estimate.¹⁰ Using the overall average, rather than the average effect by outcome domain, is an acceptable way to describe impacts of TPP programs. We concluded this after we conducted an inferential analysis to test whether the average effect sizes differed by outcome domain. We found that there was not a statistically significant difference in these averages ($p = .47$). This means that, while the average effect sizes vary slightly by outcome, it is unlikely that they represent true differences. Tabular presentations of this information, with additional details, are presented in the Technical Appendix.

Exhibit 3. Average effect sizes of TPP programs overall and by outcome domain



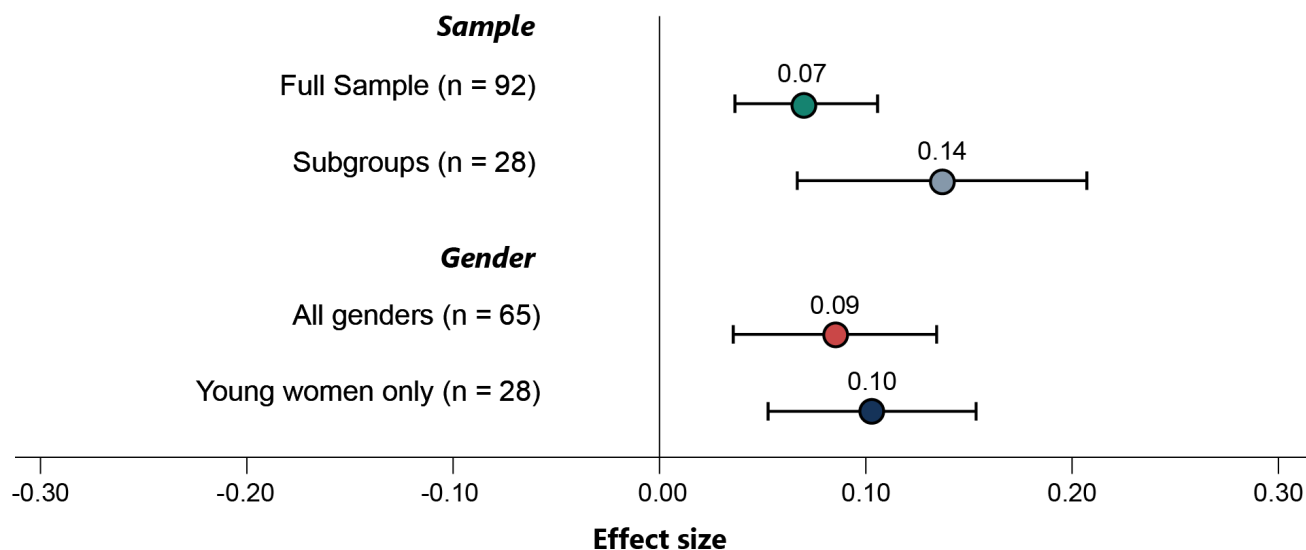
Note: n refers to the number of studies contributing to the averages. The figure presents the average effect as a dot, and a 95 percent confidence interval as lines around the dot. (Wider confidence intervals represent greater noise or uncertainty in the estimated average.) When the line does not cross the vertical effect size line at 0, the average effect is statistically significantly different from zero.

Average effect size of TPP programs tended to be larger for subgroups than the full sample, though the difference is not statistically significant

There was little variability in effect sizes observed across outcome domains, but effect sizes tended to vary across sample characteristics (Exhibit 4). Impact estimates tended to be larger in analyses examining subgroups (effect size = 0.14, $p < .01$) than in full-sample analyses (effect size = 0.07, $p < .01$), though, this difference was not statistically significantly different from zero ($p = .08$). We also examined whether impacts varied depending on the gender of participants in the study sample: there were no statistically significant differences ($p = .61$), with an average effect size of 0.09 ($p < .01$) for programs evaluated with samples that included youth of all genders and an average effect size of 0.10 ($p < .01$) for programs evaluated with samples that included young women only. Due to small sample sizes, average impacts could not be reliably calculated for programs that were evaluated with samples that included young men only or where the majority of youth reported other genders.

¹⁰ The overall average effect is the most precise estimate because it is based on all 618 available findings, whereas the findings by domain used between 30 and 244 estimates for each estimate.

Exhibit 4. Average effect sizes of TPP programs by sample characteristic



Note: n refers to the number of studies contributing to the averages. The figure presents the average effect as a dot, and a 95 percent confidence interval as lines around the dot. (Wider confidence intervals represent greater noise or uncertainty in the estimated average.) When the line does not cross the vertical effect size line at 0, the average effect is statistically significantly different from zero. TPPER-eligible subgroups are defined by either gender or sexual experience at baseline.

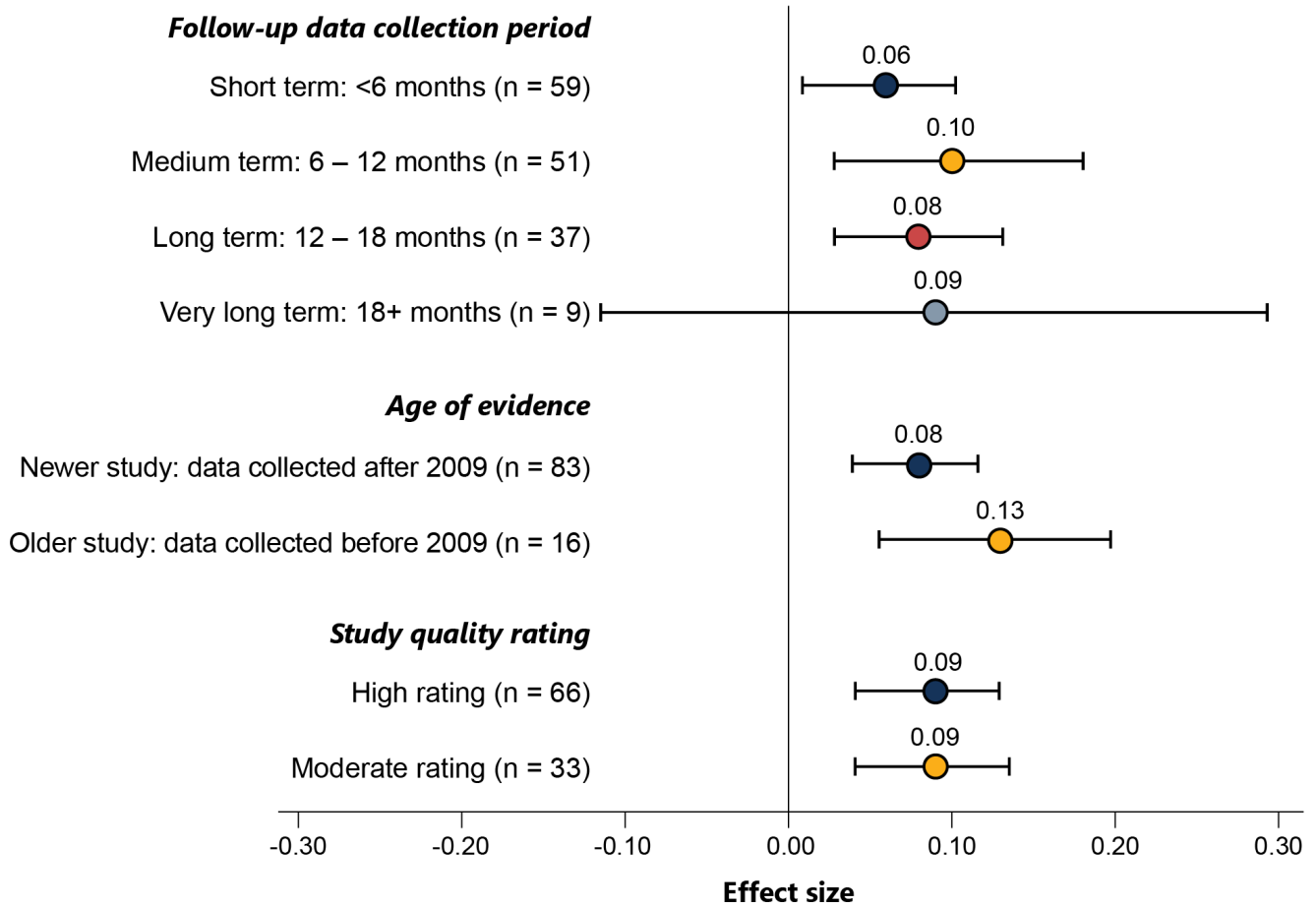
Average effect size of TPP programs was positive but did not vary significantly for all examined study characteristics, including age of evidence, study quality rating, and follow-up data collection periods

In addition to some differences in findings by sample characteristics, the magnitude of impacts varied slightly depending on characteristics of the study. Exhibit 5 illustrates small differences in the timing of follow-up data collection. There are positive and statistically significant average program effects for the follow-up periods less than six months, six to 12 months, and 12 to 18 months after the program ended. The average program effect for follow-ups in the very long term (18 or more months after the program ended) was not statistically significant. We found that the difference in average program effects across follow-up periods was not significantly different from zero ($p = .71$).

We also observed some small differences in impact findings depending on the characteristics of the study (Exhibit 5). Average program effects were larger when study data were collected before 2009 (effect size = 0.13, $p < .01$) than when study data were collected in 2009 or later (effect size = 0.08, $p < .01$), but this difference was not significantly different from zero ($p = .17$). This difference in magnitude may have resulted from (1) a large influx of replication studies of evidence-based programs starting in 2010, where the observed impacts tended to be smaller than those observed in the original studies (Goesling 2015; Valentine et al. 2011; Cole 2014), and (2) the general decrease in the rates of pregnancy and other risk behaviors among youth over time (Osterman et al. 2023).

Average program effects by high- or moderate-quality study ratings were statistically significantly different from zero (effect size = 0.09, $p < .01$ for both average estimated effects), but they are not statistically different from each other ($p = .95$).

Exhibit 5. Average effect sizes of TPP programs, by follow-up period and other study characteristics

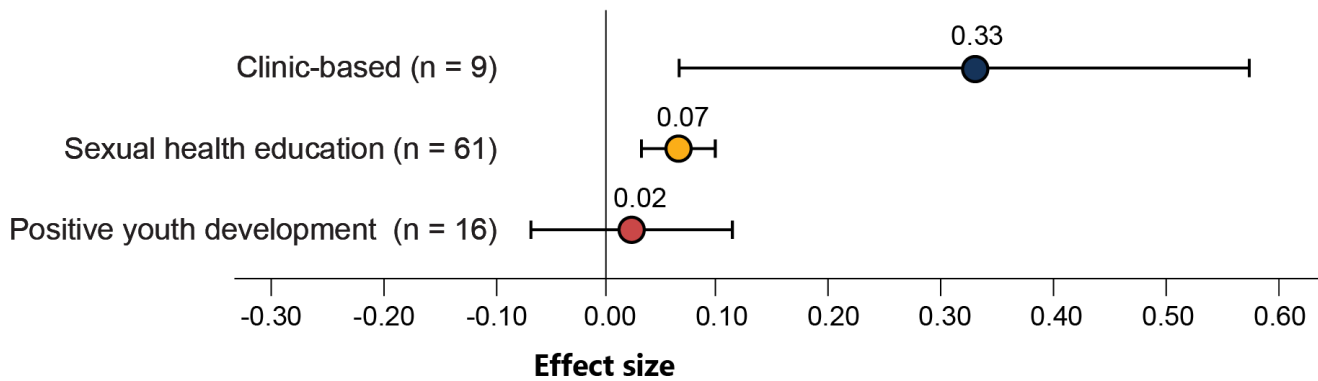


Note: n refers to the number of studies contributing to the averages. The figure presents the average effect as a dot, and a 95 percent confidence interval as lines around the dot. (Wider confidence intervals represent greater noise or uncertainty in the estimated average.) When the line does not cross the vertical effect size line at 0, the average effect is statistically significantly different from zero.

Average effect size of TPP programs varies by program type, with clinic-based programs having the largest average effect size

Finally, we calculated average program effects by program type (Exhibit 6) and found that the average program effect for clinic-based and sexual health education programs were statistically significantly different from zero (effect size = 0.33, $p < .05$ and effect size = 0.07, $p < .01$, respectively). The average program effect for positive youth development programs was not statistically significant (effect size = 0.02, $p = .57$). Additionally, we found that the average program effects by program type are statistically significantly different from each other ($p = .04$). Therefore, we conclude that it is appropriate to discuss the program effects separately for these three program types. Impacts for sexual risk avoidance and healthy relationship programs could not be credibly estimated due to small sample sizes and are not presented here.

Exhibit 6. Average effect sizes of TPP programs, by program type



Note: n refers to the number of studies contributing to the averages. The figure presents the average effect as a dot, and a 95 percent confidence interval as lines around the dot. (Wider confidence intervals represent greater noise or uncertainty in the estimated average.) When the line does not cross the vertical effect size line at 0, the average effect is statistically significantly different from zero.

Additional findings, including descriptive statistics for all results presented above, along with additional meta-analytic averages for full samples and subgroups, are available in the Technical Appendix.

CONCLUSION

TPPER systematically reviews evidence about TPP programs and highlights those programs that research has shown to be effective. At the time this brief was published, 48 programs had evidence of effectiveness by virtue of having at least one favorable, statistically significant impact on a sexual behavior outcome from a high- or moderate-quality study (Forrester et al. 2024). However, there are 31 other programs reviewed by TPPER that have eligible, credible evidence from a high- or moderate-quality study, but do not show evidence of effectiveness. This meta-analysis incorporates the evidence from 99 studies of 79 TPP programs reviewed by the TPPER to describe the effectiveness of TPP programs, overall.

The meta-analysis of the TPPER effect size data presents useful information to the TPP field: when we examine the full body of recent, credible evidence about TPP programs, we find that TPP programs as a whole are effective, and have an average effect size of 0.09, which can be conceptualized as a reduction in prevalence from 50 percent to 45.7 percent in a risky sexual behavior. This is a small but meaningful effect, given the breadth of TPP implementation and the number of youth receiving services nationwide.

TPP programs have statistically significant, favorable effects on STI/HIV, contraceptive use and sexual activity outcomes. The findings from this meta-analysis highlight both the promise of TPP programming and the gaps where more research—and potentially new programs—are needed. In particular, research examining effectiveness of programs designed specifically for young men, examining impacts on pregnancy outcomes, and examining healthy relationship and sexual risk avoidance programming; these were areas where findings were sparse and average impacts could either not be reliably estimated or where the results tended to be small and not significantly different from zero.

Comparison of current findings to previous TPP meta-analysis study

We were able to replicate the main takeaway findings of Juras et al. (2022) by restricting our analysis of TPPER data to the subset of studies that were included in that meta-analysis. When we estimated impacts for the same studies, we also observed small, non-significant average effects of TPP programs for all five outcome domains. This finding gave us confidence that the TPPER effect size data were telling a story consistent with that of another established meta-analysis in this field.

Importantly, a substantive difference in the current TPPER meta-analysis is the inclusion of roughly twice as many studies (99) as those in the Juras et al. (2022) meta-analysis, including several from an earlier point in time. As noted, the average effects from studies with data collection before 2009 tended to be larger (effect size = 0.13) than those from studies with data collection in or after 2009 (effect size = 0.08). This may have been a contributing reason why the findings across these two analyses differed. (Additional discussion is in the Technical Appendix.)

Implications, limitations, and future directions

The findings of this meta-analysis can be particularly informative for researchers. The findings reported here (and in the Technical Appendix) provide estimates of the average effects of a large body of TPP programs overall, by outcome domain, subgroup, follow-up time-period, and program type. Information about average effects (and variation around the average) is critically important for TPP researchers conducting power or minimum detectable impact (MDI) calculations. These findings can anchor MDI calculations for a given study against a plausible range of effects observed in the field (for a given outcome or program type, etc.). In addition, these average effect size statistics (and variation in observed effects around the averages) are particularly useful for calculating the probability that a given impact estimate represents a favorable effect under a Bayesian interpretation—potentially a more useful piece of information than relying on whether a finding is statistically significantly different from zero or not (Deke and Finucane 2023; Deke et al. 2022). More information on this is available in the appendix.

While there are substantive contributions from this meta-analysis, it also has limitations. We used a large collection of effect size data from 99 effectiveness studies, but there were missing effect size and standard error data from some of the findings in the included studies, and thus, those findings were not included in the analyses. Furthermore, there were 10 additional studies where all of the findings had to be eliminated due to missing effect size or standard error information. As a result, the findings reported here represent most, but not all, of the reviewed moderate- and high-quality evidence currently deemed to be eligible by TPPER.

This brief highlights the feasibility of doing meta-analytic approaches for pooling evidence about TPP programs. This meta-analysis also reveals opportunities for further exploration. One possible approach would be to examine additional features of the study samples to understand if TPP programs are particularly effective for youth in rural or urban settings or for certain racial or ethnic groups. Another important approach would be to better understand the roles of components (such as content, activities, formats) of the programs, including the roles they play as predictors of impacts (Forrester and Cole 2023). In 2023, TPPER began assembling detailed component profiles of all evidence-based programs. Going forward, meta-regression research could leverage these detailed data to better understand the ingredients of TPP programs that appear to be driving outcomes.

REFERENCES

1. Forrester, E., Rooney, C., Musaddiq, T., Knab, J., Lugo-Gil, J., Cole, R., Ouellette, L., Denk, E., Henke, J., Hollie, B., Hulse, L., and Manzer, J. "Updated Findings from the HHS Teen Pregnancy Prevention Evidence Review- May 2023." Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, 2024.
https://youth.gov/sites/default/files/2024-09/TPPER-Findings-Technical-Brief-2024_1.pdf.
2. Juras, R., M. Kelsey, K. Steinka-Fry, M. Lipsey, J. Layzer, and E. Tanner-Smith. "Meta-Analysis of Federally Funded Adolescent Pregnancy Prevention Program Evaluations." *Prevention Science*, vol. 23, 2022, pp. 1169–1195. <https://doi.org/10.1007/s11121-022-01405-0>.
3. Moreno, Lorenzo, and Russell Cole. "Calculating Minimum Detectable Impacts in Teen Pregnancy Prevention Impact Evaluations." Evaluation Technical Assistance Brief no. 7. Submitted to the Office of Adolescent Health and the Administration on Children, Youth, and Families Teenage Pregnancy Prevention Grantees. Mathematica, November 2023.
4. Deke, J., and M. Finucane. "Bayesian Interpretation of Estimates." Webinar and resources provided to Office of Population Affairs TPP Tier 2 Phase II Grantees, September 2022. Resources available at <https://rhntc.org/resources/bayesian-interpretation-estimates-webinar>.
5. Deke, J., M. Finucane, and D. Thal. "The BASIE (BAYesian Interpretation of Estimates) Framework for Interpreting Findings from Impact Evaluations: A Practical Guide for Education Researchers." NCEE 2022-005. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2022.
6. Streke, A., E. Dench, J. Lugo-Gil, and R. Cole. "Meta-Analysis Report of Teen Pregnancy Prevention Evidence Review (TPPER) Studies: Technical Appendices." U.S. Department of Health and Human Services, Assistant Secretary for Planning and Evaluation, Office of Human Services Policy, 2024.
7. What Works Clearinghouse. "What Works Clearinghouse Procedures and Standards Handbook, Version 4.0." U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2017. This report is available on the What Works Clearinghouse website at <https://ies.ed.gov/ncee/wwc/Handbooks>.
8. Goesling, B. "Making Sense of Replication Studies: Guidance for Teen Pregnancy Prevention Researchers." ASPE Research Brief. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, May 2015.
9. Valentine, Jeffrey C., Anthony Biglan, Robert F. Boruch, Felipe González Castro, Linda M. Collins, Brian R. Flay, Sheppard Kellam, Eve K. Mościcki, and Steven P. Schinke. "Replication in Prevention Science." *Prevention Science*, vol. 12, 2011, pp. 103–117.
10. Cole, Russell. "Comprehensive Reporting of Adolescent Pregnancy Prevention Programs." *American Journal of Public Health*, vol. 106, S15–S16, <https://doi.org/10.2105/AJPH.2016.303332>.
11. Osterman, M., B. Hamilton, J. Martin, A. Driscoll, and C. Valenzuela. "Births: Final Data for 2021." *National Vital Statistics Report*, vol 72, no. 1, 2023.
<https://www.cdc.gov/nchs/data/nvsr/nvsr72/nvsr72-01.pdf>.
12. Forrester, Emily, and R. Cole. "Core Components of Teen Pregnancy Prevention Programs." OPA Brief. Submitted to the Office of Population Affairs. Mathematica, April 2023.
13. What Works Clearinghouse. "What Works Clearinghouse Procedures and Standards Handbook, Version 5.0." U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2020.
14. Rotz, Dana, Emily Sama-Miller, and Paul Burkander. "Protocol for the Pathways to Work Evidence Clearinghouse: Methods and Standards." OPRE Report # 2020-44, Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2020.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of the Assistant Secretary for Planning and Evaluation

200 Independence Avenue SW, Mailstop 447D
Washington, D.C. 20201

For more ASPE briefs and other publications, visit:

aspe.hhs.gov/reports



ABOUT THE AUTHORS

Russell Cole and *Julieta Lugo-Gil* are principal researchers at Mathematica. *Andrei Streke* is a researcher at Mathematica.

SUGGESTED CITATION

Cole, R. Lugo-Gil, J., Streke, A. The effectiveness of Teen Pregnancy Prevention programs – a meta-analysis. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. December 2024.

COPYRIGHT INFORMATION

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

DISCLOSURE

This communication was printed, published, or produced and disseminated at U.S. taxpayer expense.

Subscribe to ASPE mailing list to receive email updates on new publications:

<https://list.nih.gov/cgi-bin/wa.exe?SUBED1=ASPE-HEALTH-POLICY&A=1>

For general questions or general information about ASPE:

aspe.hhs.gov/about