impact of adolescent pregnancy prevention programs for younger adolescents is often limited to a small portion of the population already engaging in the outcomes being evaluated. Finding salient, proximal precursors that are prevalent in a larger percentage of the population receiving the interventions, more predictive of sexual risk, and sensitive to change could help address this challenge. Theory-based psychosocial outcomes are prevalent, sensitive to change and serve as important secondary outcomes of prevention programs, but because they are not always predictive of sexual behavior we need more persuasive alternatives. We advocate exploring a set of strongly predictive, more proximal precursors that fall between high-prevalence, theory-based psychosocial variables and low-prevalence sexual behaviors; we also support looking beyond individual-level outcomes. AJPH

Karin K. Coyle, PhD
Jill R. Glassman, PhD

CONTRIBUTORS
All authors contributed equally to this editorial.

ACKNOWLEDGMENTS
The authors would like to thank Pamela Anderson, PhD, for her review and comments on an earlier draft of this editorial.

REFERENCES

1. Potter SC, Coyle KK, Glassman JR, Kershner S, Prince MS. *It's Your Game…. Keep it Real in South Carolina*: a group randomized trial evaluating the replication of an evidence-based adolescent pregnancy and sexually transmitted infection prevention program. *Am J Public Health*. 2016;106(suppl 1):S60–S69.

2. Coyle KK, Anderson PM, Franks HM, Glassman JR, Walker JD, Charles VE. Romantic relationships: an important context for adolescent HIV/STI and pregnancy prevention programs. *J Sex Educ*. 2014;14(5):582–596.

3. Glassman JR, Franks HM, Baumler ER, Coyle KK. Mediation analysis of an adolescent HIV/STI/pregnancy prevention intervention. *J Sex Educ*. 2014;14(5):497–509.

4. Ballonoff Suleiman A, Brindis CD. Adolescent school-based sex education: using developmental neuroscience to guide new directions for policy and practice. *Sex Res Soc Policy*. 2014;11(2):137–152.

5. O'Sullivan LF, Cheng MM, Harris KM, Brooks-Gunn J. I wanna hold your hand: the progression of social, romantic and sexual events in adolescent relationships. *Perspect Sex Reprod Health*. 2007;39(2):100–107.

6. Hennessy M, Bleakley A, Fishbein M, Jordan A. Validating an index of adolescent sexual behavior using psychosocial theory and social trait correlates. *AIDS Behav*. 2008;12(2):321–331.

7. Peper JS, Dahl RE. The teenage brain: surging hormones—brain behavior interactions during puberty. *Curr Dir Psychol Sci*. 2013;22(2):134–139.

8. Connolly J, Craig W, Goldberg A, Pepler D. Mixed-gender groups, dating, and romantic relationships in early adolescence. *J Res Adolesc*. 2004;14(2):185–207.

Establishing an Evaluation Technical Assistance Contract to Support Studies in Meeting the US Department of Health and Human Services Evidence Standards

This special issue highlights the results of the Office of Adolescent Health's (OAH) substantial investment in rigorous evaluations of teen pregnancy prevention (TPP) programs. Through a two-tiered funding strategy, OAH procured cooperative agreements with 94 grantees to replicate programs deemed evidence-based by the US Department of Health and Human Services' (HHS) TPP evidence review (tier 1) or to implement promising and innovative TPP programs that did not yet have evidence of effectiveness (tier 2). In addition, the Family and Youth Services Bureau, under the Administration for Children and Families (ACF) at HHS, also funded 13 cooperative agreements to implement promising programs through the Personal Responsibility Education Program Innovative Strategies program. A subset of the cooperative agreements required the grantees to evaluate the effectiveness of their funded programs through random assignment or quasi-experimental impact studies led by independent evaluators. The goal of this investment in evaluation was to infuse the field with dozens of new, internally valid studies whose evidence would meet the rigorous research standards established by the HHS TPP evidence review[1] and would inform the field of public health. These new findings would be used to further understand the effectiveness of evidence-based programs when implemented in different contexts and for different populations, and to potentially identify new, effective programs.

OAH's investment is part of a larger federal effort to use and create evidence through tiered-evidence grant programs.[2] As the government encourages and incentivizes rigorous evaluations,[3] some large-scale federal grant programs provide evaluation technical assistance (TA) to their grantee-led evaluations, including the Investing in Innovation Grants (I3), administered by the US Department of Education; and the Workforce Innovation Fund Grants, administered by the Employment and Training Administration under the US Department of Labor. To support grantees in producing credible evidence of program effectiveness, OAH (with support from ACF) funded Mathematica Policy Research and its subcontractors to be the evaluation technical assistance

ABOUT THE AUTHORS
Russell P. Cole and Susan Goerlich Zief are Senior Researchers at Mathematica Policy Research, Princeton, NJ. Jean Knab is an Associate Director of Human Services Research at Mathematica Policy Research.

Correspondence should be sent to Jean Knab, Mathematica Policy Research, PO Box 2393, Princeton, NJ 08543-2393 (e-mail: jknab@mathematica-mpr.com). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This editorial was accepted June 23, 2016.
doi: 10.2105/AJPH.2016.303359

contractor for the OAH and ACF grantee-led rigorous evaluations that were not part of federally led evaluations, with oversight by the OAH evaluation specialist.

Most of the federal evaluation TA efforts, OAH's included, are structured around meeting a very specific goal: that completed studies meet a particular set of evidence standards for their field. For OAH, this is the HHS TPP evidence review standards, described below. These TA efforts have broader goals, as well. According to Gibbs et al.,[4] evaluation TA on individual studies can lead more broadly to improved capacity in the field to conduct evaluations. For example, the OAH evaluation TA effort can help the grantee organizations and their partner evaluators to lead and produce future evaluations that will meet rigorous evidence review standards, regardless of whether OAH is supporting them. Increasing the rigor of research in the field should lead to a better understanding of what works in field of public health.

HHS EVIDENCE STANDARDS

The HHS TPP Evidence Review assesses the credibility of the evidence of programs aiming to reduce adolescent pregnancies, sexually transmitted infections, and sexual risk behaviors. The evidence review's activity is conducted in two steps. First, the evidence review systematically assesses the *quality of the evidence* from a study, and second, for the subset of studies that are deemed to provide credible evidence, the review describes the *effectiveness* of the program described in the study.

The assessment of the evidence places studies into a high-, moderate-, or low-quality evidence category. This categorical assessment helps differentiate the trustworthiness, or internal validity, of the evidence generated from a study. The review process examines features of the study design and evaluation implementation (for example, well-implemented randomized controlled trials are eligible for the high-quality evidence rating, but quasi-experimental designs are only eligible to receive a moderate rating because of the potential threats to internal validity associated with the design). The review process also takes into account other threats to internal validity, such as sample attrition (nonresponse), equivalence at baseline, and factors that confound an assessment of program impacts. Illustrative examples of study categorization are as follows:

- *High quality:* Randomized controlled trials with low levels of sample attrition, and statistical controls for any baseline nonequivalence.
- *Moderate quality:* Randomized controlled trials with high attrition or quasi-experimental designs, and the study demonstrates baseline equivalence.
- *Low quality:* Randomized controlled trials with high attrition or quasi-experimental designs, and the study does not demonstrate baseline equivalence.

Studies with a high or moderate rating are considered as having internally valid evidence and are eligible for an assessment of program effectiveness. If a study with a high or moderate evidence rating shows a statistically significant, favorable impact of a program on a sexual behavior outcome, the program is labeled as having *evidence of effectiveness* by the evidence review. A current list of the 44 programs deemed to be effective by the evidence review, as well as a fuller description of the study eligibility and review criteria, is available at http://tppevidencereview.aspe.hhs.gov.

Notably, similar to other federal systematic review efforts, the HHS TPP evidence review focuses on the internal validity of the evidence: whether the observed impact can be causally attributed to the program being tested. The evidence review does not assess the extent to which evidence from any one study is generalizable to other populations and settings. Most of the studies conducted to date have been implemented in single geographic areas, and the samples are not considered representative of a larger population. In reporting findings, the HHS TPP evidence review does describe the population and setting of each study, allowing users of the evidence review findings to identify relevant populations and settings.

LEARNING FROM A LARGE-SCALE TA EFFORT

Two related editorials describe the role that evaluation TA played in assisting grantee efforts to meet the HHS evidence standards. Specifically, Zief et al.[5] describe how the evaluation TA was structured to support over 40 grantee evaluations (both OAH and ACF grants) throughout the funding period, from initial study design through final reports, in an effort to enable studies to meet evidence standards. And Knab et al.[6] describe the primary challenges germane to this particular evaluation TA effort, how OAH and the evaluation TA team addressed the challenges, and the implications and lessons learned for future evaluation TA efforts. As a collection, these three editorials, along with Margolis and Roper,[7] provide some context for how funders or stakeholders can support and maintain grantee-led evaluation efforts through evaluation TA.

While this large-scale effort was costly for OAH, by most accounts the TA was a success. The final reports are credible, internally valid presentations of the effects of the programs, although, several of them have limited external validity (generalizability). All of the final evaluation reports submitted to OAH are expected to meet HHS evidence standards and be rated as having high- or moderate-quality evidence. The evaluation capacity in the field has been strengthened by these public health grantees and evaluators' participation in this effort. AJPH

Russell P. Cole, PhD
Susan Goerlich Zief, PhD
Jean Knab, PhD

CONTRIBUTORS
All authors contributed equally to this editorial.

ACKNOWLEDGMENTS
This work was conducted under a contract (HHSP233201300416G) with the Office of Adolescent Health, within the Department of Health and Human Services (HHS).

REFERENCES

1. US Department of Health and Human Services. Identifying programs that impact teen pregnancy, sexually transmitted infections, and associated sexual risk behaviors. Review Protocol, Version 4. Available at: http://tppevidencereview.aspe.hhs.gov. Accessed February 22, 2016.

2. Council of Economic Advisers. *Economic Report of the President*. Washington, DC: Council of Economic Advisers; 2014.

3. General Accounting Office. Program evaluation: strategies to facilitate agencies' use of evaluation in program management and policy making. GAO Publication No. 13–570. Washington, DC: Government Printing Office; June 2013.

4. Gibbs DA, Hawkins SR, Clinton-Sherrod AM, Noonan RK. Empowering programs with evaluation technical

assistance: outcomes and lessons learned. *Health Promot Pract.* 2009;10(1): 38S–44S.

5. Zief SG, Knab J, Cole RP. A framework for evaluation technical assistance. *Am J Public Health.* 2016;106(suppl 1): S24–S26.

6. Knab J, Cole RP, Zief SG. Challenges and lessons learned from providing large-scale evaluation technical assistance to build the adolescent pregnancy evidence base. *Am J Public Health.* 2016;(suppl 1): S26–S28.

7. Margolis AL, Roper A. Practical experience from the Office of Adolescent Health's large scale implementation of an evidence-based teen pregnancy prevention program. *J Adolesc Health.* 2014;54: S10–S14.

A Framework for Evaluation Technical Assistance

The primary goal of the evaluation technical assistance (TA) was to help the grantees and their evaluators produce rigorous evidence of program effectiveness that would meet the US Department of Health and Human Services (HHS) Teen Pregnancy Prevention (TPP) evidence review standards, described in Cole et al.[1] To achieve this goal, the evaluation TA team created a framework for TA that included an initial design review, ongoing monitoring and group TA, and a series of evaluation deliverables designed to serve as stepping stones toward the final evaluation report.

This editorial is the second in a series of related opinion pieces. As noted above, the first editorial described the context of the evaluation TA contract, and outlined the HHS evidence standards that were used as an operational goal for a high-quality evaluation. This second editorial describes the TA framework.The final editorial provides a summary of the challenges and lessons learned from the evaluation TA project, and offers insight into the ways that evaluation TA has been improved for future grantee funded evaluations.[2]

DESIGN REVIEWS (YEAR 1)

The first step for funded grantees was receiving approval of their proposed designs. We reviewed the planned evaluation designs to determine whether they could potentially meet the HHS TPP evidence review standards, provided the evaluations were well implemented over the subsequent five years, and be sufficiently powered to detect statistically significant impacts. Unlike the evidence review, which assessed completed studies, the design review assessed evaluation *plans* to determine the likelihood that aspects of the design and implementation would not introduce sample bias and the final evidence would meet the evidence review standards. We took the evidence review standards and worked backward to be sure the proposed designs and data collection plans were aligned with the final standards.

The team used a design review template to systematically review each proposed design, assessing features including, but not limited to the (random) assignment procedure, the approach for recruiting and consenting sample members, and plans for sample retention and data collection. These features were assessed to determine whether they were potential threats to the validity of the study. For instance, when we encountered scenarios that might compromise the integrity of the random assignment design (for example, moving youths from the condition to which they were randomly assigned to a different condition), we worked to the extent possible with the grantee and evaluator to reduce potential threats to the study meeting standards. In addition, depending on feasibility and logistical and resource constraints, we attempted to improve the statistical power of each study design, if needed, to increase the likelihood of the study showing statistically significant impacts.

During the design review phase, the evaluation TA team provided feedback to grantees through written reviews, telephone calls, and in-person discussions. The Office of Adolescent Health (OAH) decided when to approve each design after receiving input from the evaluation TA team, at which point the evaluation TA team shifted gears toward monitoring the quality of the ongoing evaluations.

ONGOING MONITORING AND GROUP TA (YEARS 2–4)

The second phase of evaluation TA was monitoring the implementation of the evaluation and identifying solutions to real-world problems that affect nearly all rigorous impact evaluations once they go into the field. The evaluation TA team conducted ongoing monitoring through individual TA and also provided group TA for common issues faced across the evaluations.

- **Ongoing monitoring:** Each grantee and their evaluator worked closely with a TA liaison, who served as the point person for the broader evaluation TA team. In conjunction with the project officer, the TA liaison held monthly monitoring calls during the implementation period to identify risks to the integrity of the study and to brainstorm suggested solutions for addressing them. Individual TA was an important component of the TA framework, so the TA could be customized to the needs and experiences of the evaluation team.
- **Group TA:** Above and beyond the individual TA provided to each grantee and their evaluators, the evaluation TA team provided a wide array of group TA. The group TA included presentations of timely evaluation topics at grantee conferences, webinars and conference calls, and written briefs on evaluation topics, such as descriptions of the evidence review standards, methods for handling missing or inconsistent self-report

ABOUT THE AUTHORS

Susan Goerlich Zief and Russell P. Cole are Senior Researchers at Mathematica Policy Research, Princeton, NJ. Jean Knab is an Associate Director of Human Services Research at Mathematica Policy Research.

Correspondence should be sent to Jean Knab, Mathematica Policy Research, PO Box 2393, Princeton, NJ 08543-2393 (e-mail: jknab@mathematica-mpr.com). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This editorial was accepted June 23, 2016.

doi: 10.2105/AJPH.2016.303365