

CONCLUSIONS

The evaluation TA contract funded by OAH was successful in establishing a framework for a large-scale evaluation TA effort. The team developed processes and report templates that facilitated training and TA to more than 100 grantees and evaluator partners, as well as timely reporting to

OAH. The framework that was developed, with improvements made from lessons learned,² is also being used for OAH's cohort 2 and has been adopted by similar efforts outside of OAH. *AJPH*

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

Challenges and Lessons Learned From Providing Large-Scale Evaluation Technical Assistance to Build the Adolescent Pregnancy Evidence Base

This is the third editorial in a series of related opinion pieces. The first editorial provided a description of the Office of Adolescent Health's (OAH) investment in teen pregnancy prevention (TPP) programs and the evaluation technical assistance (TA) contract that was intended to support funded grantees.¹ That editorial also outlined the US Department of Health and Human Services (HHS) evidence standards that guided the evaluation TA contract. The second editorial detailed the activities conducted under the evaluation TA contract with the first cohort of funded grantees.² Implementing a large-scale TA effort was not without challenges. These challenges were necessary and expected, as the federal government looks to support large-scale grantee-led evaluations to build the evidence base in TPP. The lessons learned from these challenges laid the groundwork for a second round of evaluation TA with a second cohort of TPP grantees that began in summer 2015.

These challenges and their solutions can also inform other federally sponsored evaluation TA efforts. The following is a discussion of the major challenges in providing TA to the grantees in cohort 1. These challenges were identified through the TA activities laid out in Zief et al.,² primarily our monitoring calls and document reviews.

CHALLENGE 1: NO STANDARDIZED APPROACH

Federal efforts to assess impact evaluations for the credibility of their causal inferences began with the Department of Education's What Works Clearinghouse in 2004 and have spread throughout other agencies and offices, such as HHS and the Department of Labor. These agencies have each developed similar sets of standards that are used to systematically and consistently assess the internal validity of completed studies in

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. Cole RP, Zief SG, Knab J. Establishing an evaluation technical assistance contract

to support studies in meeting the HHS evidence standards. *Am J Public Health*. 2016;106(suppl 1):S22–S24.

2. 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; 106(suppl 1):S26–S28.

products that could inform the broader field and future grantees and evaluators. The contract produced research briefs on planning evaluations designed to meet evidence standards, coping with missing data and clustering in randomized controlled trials, baseline equivalence and matching, attrition in randomized controlled trials, best practices for school and district recruitment, and a primer on the evidence standards. The dissemination materials, available on the OAH Web site,³ supplied a road map for the second cohort of grantees that did not exist for the first cohort.

CHALLENGE 2: NOT THE SOLE BENCHMARK

The HHS evidence review assesses the internal validity of the impact findings to document the extent to which the results are credible and could therefore guide policy and program decisions. However, for most policymakers and researchers, the key question of interest is

their field. But no publicly available road maps existed within HHS, or any other federal agency, for designing and implementing an evaluation that has a good chance of resulting in a completed study that meets evidence review standards. Because many grantees and evaluators were initially unaware or only partially aware of the standards, they were not considering the evidence standards when designing their evaluations. So this contract focused on building a process (the framework noted in Zief et al.²) for training the grantees on the evidence standards and best practices for meeting those standards. This was accomplished through providing formal and informal training during the TA process and developing written

ABOUT THE AUTHORS

Jean Knab is an Associate Director of Human Services Research at Mathematica Policy Research, Princeton, NJ. Russell P. Cole and Susan Goerlich Zief are Senior Researchers 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.2016.303358

whether the program had an effect that was statistically significant, which is in part based on the study's power and the effective contrast in services across the treatment and control groups.

When the evaluation TA contractor began design plan reviews, the team identified numerous studies with low statistical power or a small-expected programmatic contrast between the two groups. Upon consultation with OAH, the evaluation TA team made recommendations for numerous studies to improve statistical power and contrast. For example, some studies proposed a small sample (e.g., less than 500) but planned multiple (three or more) follow-up surveys. Eliminating one follow-up survey sometimes provided the resources needed to enroll a larger sample. Others proposed larger samples but had a control group receiving very similar services as the treatment group. Therefore, such studies were unlikely to observe large differences in participant outcomes, given the small differences in the effective contrast between treatment and control groups. More often than not, recommendations to improve statistical power and the contrast between the two groups had budget implications that could not always be offset by other design modifications, or they were infeasible to implement given the limitations of the evaluation settings. As a result, a small number of studies were unable to accommodate the recommended improvements and moved forward with expected weak contrasts between the two groups and low power to detect statistically significant impacts.

As OAH prepared their funding announcement for the

second cohort, the TA contract prepared a research brief and power calculator targeted toward research in this field. The goal was to encourage applicants to be more thoughtful and critical when determining the optimal sample size and contrast for the design. Study power and contrast were also focal points of the early reviews of applicant and funded design plans for the second cohort.

CHALLENGE 3: LATE START TO TA

One reason that the evaluation TA team was limited in its ability to improve some of the evaluations is that the TA contract was funded after the grantees were awarded. The Office of Adolescent Health was a new federal office funded in 2010, and it had to move quickly to launch the grant program, the evaluation TA contract, and a related performance measures contract during fiscal year 2010.⁴ The evaluation TA activities started several months after the grant awards were announced. Because of this timeline, grantees and their evaluators were asked to rethink and refine their original evaluation and program plans nearly a year after they received their grant award. Understandably, this led to quite a bit of confusion with and frustration over a newly evolving set of expectations and requirements. For some grantees, the process and timeline led to delays in beginning the programming and evaluation. For the TA team, the late start meant that budgets were set and some parameters were nonnegotiable. Needless to say, relationships between all involved parties

were at times strained during this first year.

OAH made several adjustments to the process before releasing the funding opportunity announcement for the second cohort of grantees funded in 2015. First, the evaluation TA contractor was funded prior to the awards and would therefore be available to provide evaluation TA support before and immediately upon grant award. Second, the top-scoring evaluation applications were reviewed specifically to identify any designs with a low probability of being successful. Finally, the expectations and requirements for the evaluations and work with the evaluation TA contractor was disclosed in the funding opportunity announcement and again at the time of award, preparing grantees and evaluators for evaluation TA involvement. These changes have resulted in faster approvals for the second cohort grantee designs.

CONCLUSIONS

Despite the challenges, evaluation TA improved the quality of the completed evaluations. Ironically, the TA bolstered the rigor of many studies that unfortunately do not show statistically significant program impacts; several of these studies would not have produced credible or publishable evidence without this TA (see Farb and Margolis⁵ and Cole⁶). As a result, these no-findings evaluations would never have had their evidence published or been deemed credible, contributing to what Rosenthal calls the “file-drawer problem”—though, the registration of the evaluations may have

ameliorated this problem somewhat.⁷ That said, it is important to add credible evidence to the field regardless of the direction and statistical significance of the impact estimate. Understanding that some programs have nonsignificant findings, at least in some contexts or for some populations, is an important contribution to the evidence base, particularly when some of these programs have been shown to improve participant outcomes in other settings or with different populations. Furthermore, observing substantively large, but nonsignificant impacts in an underpowered study may highlight an opportunity to conduct an additional evaluation for a potentially promising program with greater attention paid to the study design and implementation. **AJPH**

Jean Knab, PhD
Russell P. Cole, PhD
Susan Goerlich Zief, 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. Cole RP, Zief SG, Knab J. Establishing an evaluation technical assistance contract to support studies in meeting the HHS evidence standards. *Am J Public Health.* 2016;106(10):S22–S24.
2. Zief SG, Knab J, Cole RP. A framework for evaluation technical assistance. *Am J Public Health.* 2016;106(suppl 1):S24–S26.
3. US Department of Health and Human Services. Office of Adolescent Health. Evaluation training & technical assistance (TA). Available at <http://www.hhs.gov/ash/oah/oah-initiatives/evaluation/ta.html>. Accessed August 24, 2016.
4. 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):S3–S9.

5. Margolis AL. The Teen Pregnancy Prevention Program (2010–2015): synthesis of impact findings. *Am J Public Health*. 2016;106(suppl 1):S9–S15.

6. Cole RP. Comprehensive reporting of adolescent pregnancy prevention programs. *Am J Public Health*. 2016;106(suppl 1):S15–S16.

7. Rosenthal R. The “File Drawer Problem” and tolerance for null results. *Psychol Bull*. 1979;86(3):638–641.

Adolescent Pregnancy Prevention Programs and Research: A Time To Revisit Theory

Those of us engaged in the study of the effectiveness of adolescent pregnancy prevention interventions under the Office of Adolescent Health funding have dedicated time and effort to ensure the technical quality of these investigations; we have applied rigorous methods and adhered to careful reporting standards so that the estimates from our randomized trials or high-quality quasi-experimental studies can have a credible causal interpretation.

As we seek to interpret the results from this research, individually and cumulatively, it is an appropriate time to critically revisit the ideas—the theories—that ostensibly form the basis of the programs we are studying. While behavioral outcomes are rightly the focus of the Department of Health and Human Services’ evidence review and Office of Adolescent Health’s Teen Pregnancy Prevention program, understanding why these programs influence youths (or fail to do so) requires a shift of focus to the intervention’s logic model.

INCONSISTENT EVIDENCE

When we consider the evidence on programs that aim to reduce adolescent pregnancy, sexually transmitted infections, and sexual risk behaviors, we are confronted with a picture that

is puzzling. Some studies find interventions produce evidence of behavior change while others do not.¹ This variability persists within and across specific programs and time. Making sense of this puzzle requires a critical investigation of the posited processes by which the interventions are hypothesized to effect change, the application of these theories by researchers and developers, and the alternative theories that may help augment these approaches.

For instance, the theory of planned behavior and social cognitive theory (and variants of these) are among the most commonly employed theories in pregnancy and sexually transmitted infection/HIV prevention programming.² These theories have been used to predict a broad range of behaviors in correlational studies. Results from causal analyses, however, have been uneven. Interventions based on these theories may influence necessary mediating variables, but fail to effect change in the desired behavioral outcomes.³ They may also demonstrate positive impacts on behavior, but when they do, the observed effects tend to be modest, inconsistent, or diminish over time.^{4,5}

THEORY APPLICATION

Some of this may be the result of poor application of these

theories. Researchers may not be operationalizing the constructs in a way that is consistent with the theory or with other applications of the theory that exist in the literature. The mediating factors in social cognitive and planned behavior theories are latent constructs that can be difficult to measure. If validated instruments exist they may not measure the specific objects that are the target of the intervention, or they may contain too many items for a questionnaire designed primarily to measure behavioral outcomes. As a consequence, scales may be unreliable, invalid, or may fail to measure the desired construct.

Another complicating factor is that while these theories are often invoked, the connection between the theory and the programmatic components may be nebulous. Developers may identify a theoretical basis but fail to explicate or even fully consider how program components are related to key mechanisms or constructs that the theory identifies as necessary. If such an intervention fails to influence behavior, it seems spurious to infer anything about the theory itself.

THEORY LIMITATIONS

But there are also reasons to expect that the social cognitive and planned behavior theories may be limited on their own to effect lasting and meaningful risk reduction in adolescent sexual behaviors. First, at a basic level, any practicable intervention of this sort will be limited in the dosage it can provide. Whatever the programmatic exposure, it is likely modest in magnitude and perceived salience compared with the myriad other stimuli that compete for the attention of the adolescent mind each day and over time. Moreover, even if the program succeeds in changing beliefs, attitudes and intentions, the forces that brought about the change will likely diminish once the program ends. Whatever change that occurs as a result of a social cognitive intervention may therefore be expected to regress in time.

Next, what might seem like an obvious point: the social cognitive and planned behavior theories have been developed for the explanation of human behavior in general and not specifically for the reduction of sexual risk behaviors among adolescents. Given the complex factors that we currently understand are relevant to the reduction of high-risk adolescent sexual behavior, the modest or uneven impacts are not surprising. All behaviors may not be modified similarly or as robustly through intentional

ABOUT THE AUTHORS

Both authors are with The Policy & Research Group, New Orleans, LA.

Correspondence should be sent to Eric Jenner, Director of Research, The Policy & Research Group, 8434 Oak St., New Orleans, LA 70118 (e-mail: ejenner@policyandresearch.com). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints” link.

This editorial was accepted June 19, 2016.

doi: 10.2105/AJPH.2016.303333