Statement by

Evelyn Kappeler, Director, HHS Office of Adolescent Health
on
Policy-Driven Demand for Government Evaluation: Data and Capacity Needs
HHS Teen Pregnancy Prevention Program and Key Issues Related to Evaluation
before
Commission on Evidence-Based Policymaking
November 4, 2016

Good morning.

Chairwoman Abraham, Co-chair Haskins and Members of the Commission, thank you for the opportunity to present on the U.S. Department of Health and Human Services' (HHS) Office of Adolescent Health's efforts to build capacity to evaluate teen pregnancy prevention programs and to use data to improve programs.

As the Director of the HHS Office of Adolescent Health (OAH), one of my responsibilities is to lead efforts to implement and evaluate evidence-based approaches as well as to implement and test new approaches to teen pregnancy prevention. Strengthening capacity to implement evidence-based programs with quality and fidelity, evaluating those programs with rigor, and using data to inform our program work are all key priories for my office.

First funded in fiscal year (FY) 2010, the Office of Adolescent Health administers the Teen Pregnancy Prevention (TPP) Program, also known as the OAH TPP Program. The program is part of a growing movement towards evidence-based policymaking to address pressing social issues in this country, such as teen pregnancy prevention. In FY2010, the President proposed a new Teenage Pregnancy Prevention initiative to address the high teen pregnancy rates in the United States by replicating evidence-based models and testing innovative strategies. In addition, the President proposed dedicated funding to evaluate teen pregnancy prevention approaches. That year, Congress appropriated $110,000,000 for making competitive contracts and grants to public and private entities to fund medically accurate and age appropriate programs that reduce teenage pregnancy, and for the costs associated with administering and evaluating the

program. Since FY2010, through the Public Health Service Act, an additional $4,550,000 - $8,550,000 (currently $6,800,000) was made available to carry out evaluations (including longitudinal evaluations) of teenage pregnancy prevention approaches. Evaluations may be conducted on any teen pregnancy prevention approach, whether it is a federal grant or non-federally funded.

The OAH TPP Program represents a series of "firsts" – the first time the federal government established an inventory of evidence-based programs for preventing teen pregnancy; the first time the Federal government defined what rigorous evaluation meant for teen pregnancy prevention; and the first time there were funds dedicated and made available to communities to use evidence-based programs as well as test new approaches.

I will also note a report from the U.S. Government Accountability Office that examines the experience of four Federal agencies, including OAH, in identifying the potential benefits of tiered evidence-based programs, as well as the implementation challenges, and provides special examples of ways in which agencies provide technical assistance to build the capacity of funded grantees.[1]

In FY2010, OAH awarded $100 million in grant funding to a diverse set of eligible organizations, including school districts, community-based organizations, universities, and health

[1]U.S. Government Accountability Office Report, "*Tiered Evidence Grants: Opportunities Exist to Share Lessons from Early Implementation and Inform Future Federal Efforts.*" Available at http://www.gao.gov/products/GAO-16-818

departments. This funding supported grants in the form of cooperative agreements through a tiered approach that included:

- $75 million for replicating one or more evidence-based programs that were proven effective through rigorous evaluation to reduce teenage pregnancy, behavioral risk factors underlying teenage pregnancy, or other associated risk factors, (Tier 1) and
- $25 million for research and demonstration grants to develop, replicate, refine, and test additional models and innovative strategies for preventing teenage pregnancy (Tier 2).

To help identify evidence-based programs that were eligible for replication, HHS conducted an independent systematic review of the evidence base on programs to reduce teen pregnancy, sexually transmitted infections (STIs), and associated sexual risk behaviors. The first review of the teen pregnancy prevention research evidence was completed in early 2010. That review identified, assessed, and rated the rigor of program impact studies and described the strength of evidence supporting different program models. From 200 program impact studies, 28 evidence-based program models were identified. The program models vary in length and content, and include a range of approaches, including positive youth development, sex education, abstinence education, clinic-based programs, and programs for special populations including youth in the juvenile justice system, youth in alternative schools, and expectant and parenting teens.

The HHS TPP Evidence Review is updated regularly (every eighteen months to two years) and has grown since the original 28 program models identified in 2010 to now 44 different models identified. The evaluation work of the Office of Adolescent Health helped support eight new program models that meet the criteria to be added to the list of programs with evidence of effectiveness.[2] The contract is currently managed by the Office of the Assistant Secretary for Planning and Evaluation (ASPE) with funding from and in collaboration with OAH and the Administration on Children, Youth and Families (ACYF).

The findings from the first review of the research evidence were released in conjunction with the Teen Pregnancy Prevention (TPP) funding opportunity announcements (FOAs) as program models that met the evidence-standard were eligible for implementation and replication funding under the 2010 FOA. Detailed information about the evidence review and the individual program models was posted along with the FOA on the Office of Adolescent Health website to assist potential applicants and communities in reviewing the program models, accessing relevant materials, and identifying the original program developer.

From FY2010 to FY2014, OAH funded 102 grants, each were awarded between $400,000 and $4 million per year for a 5-year grant period. These grantees reached approximately 500,000 youth ages 10-19 in 39 states and the District of Columbia. Of those grants, 75 received funding

[2] Seven of the eight program models were added during the 2016 update to the HHS TPP Evidence Review. The eighth program model will be included in the 2017 update.

to replicate evidence-based programs and 27 received funding to develop and test a new or innovative approach to prevent teen pregnancy, including eight grantees funded in partnership with the Division of Reproductive Health at the Centers for Disease Control and Prevention (CDC) to implement a communitywide approach to preventing teen pregnancy. The results and lessons learned from the first cohort of TPP grantees were released July 2016, and we are now moving forward with a second five-year cohort of 84 TPP Program grant projects across the country, which are anticipated to reach 1.2 million youth in 39 States and the Marshall Islands by the end of FY2019.

Prioritizing Rigorous Evaluation in the OAH TPP Program

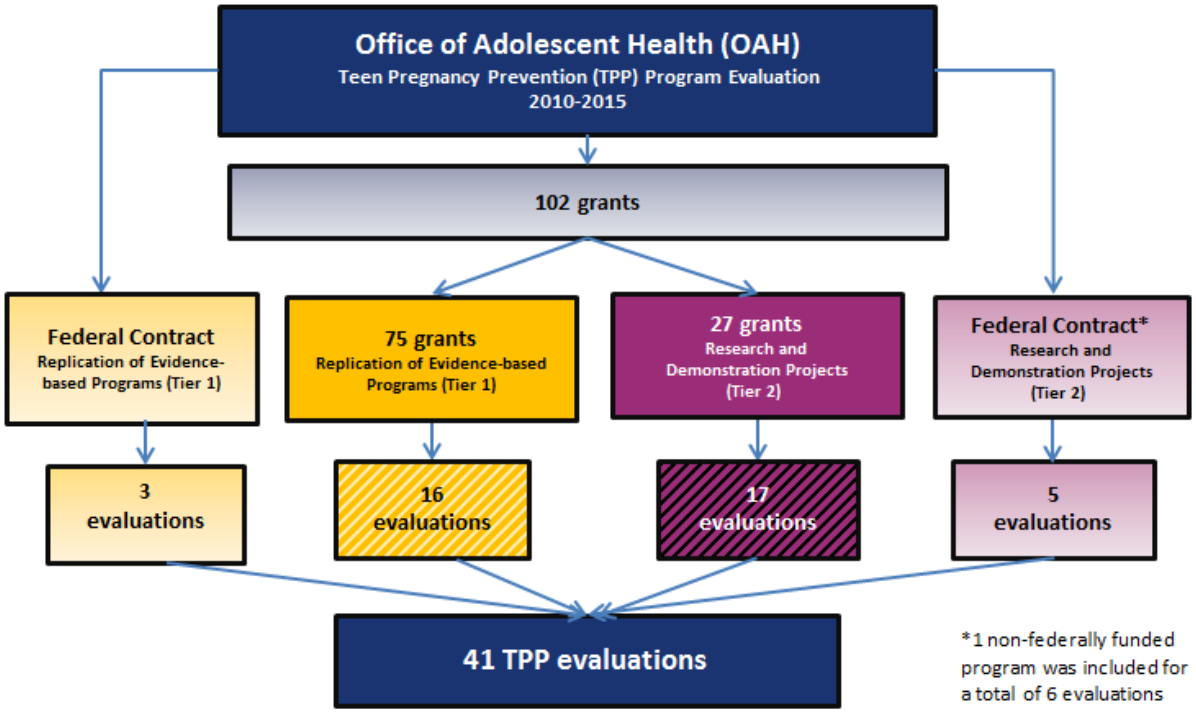
TPP Evaluation Framework FY2010-2014

OAH used a multi-pronged approach to evaluation in the first cohort of the TPP Program. That approach included the collection of performance measures by grant projects, a series of grantee-led evaluations, and two Federal-led evaluations. The purpose of this framework was to address the question of whether the replicated evidence-based teen pregnancy prevention programs and the new, innovative strategies for preventing teen pregnancy were effective.

1. In the first cohort, all 102 TPP grantees collected and reported on a common set of performance measures twice a year.
2. All Tier 1 grantees funded at $1 million per year or more (n=16) and all Tier 2 grantees rigorously evaluating a new or innovative approach (n=17) were required to conduct an independent, rigorous grantee-led evaluation of their program.
3. An additional two federal-led contracts supported 8 individual evaluations.

Figure 1 provides a graphical depiction of OAH-funded evaluation activities.

Figure 1



Performance Measures

To support grantees in having the information they needed to be successful with implementing programs, OAH collected performance measures from all grantees to assess the program's progress in achieving goals. Grantees were also encouraged to use that data to assess and improve their own program operations, ensuring continuous learning and timely feedback before the project was completed. TPP Program performance measures included measures on reach, dosage, fidelity and quality, partnerships, training, and dissemination, to assess whether grant projects were making sufficient progress toward their stated missions and serving the public interest. To measure fidelity and quality, grantees were required to collect self-report data from all program facilitators and to conduct independent observations of at least 10% of all program sessions.

Grantee-Led Evaluations

There were 33 grantee-led rigorous evaluations of TPP replication and research and demonstration grants administered by OAH from FY2010 to FY2014. OAH required that all grantee evaluations be conducted by an independent, third-party evaluator hired by the grantee and that the evaluations be designed and implemented to meet the research quality standards set for the HHS TPP Evidence Review. The benefits of rigorous local, grantee evaluations include (1) the ability to evaluate a large number of interventions in a large number of settings for less resources than would be required if done through a federally sponsored cross-site evaluation; and (2) the ability to enhance the skills of a large number of local evaluators by providing intensive evaluation technical assistance throughout the entire evaluation.

As noted in a March 2014 supplement of the *Journal of Adolescent Health*,[3] OAH found there is a set of specific conditions necessary to ensure that grantee-level evaluations maintain a high level of rigor. These include:

- Detailed criteria for what is considered a rigorous evaluation.
- A detailed description in the funding opportunity announcement of the expectations for conducting a rigorous evaluation by which applications are reviewed.

[3] Margolis, A. L., & Roper, A. Y. (2014). Practical experience from the Office of Adolescent Health's large scale implementation of an evidence-based teen pregnancy prevention program. *Journal of Adolescent Health*, *54*(3), S10-S14. Available at: http://www.jahonline.org/article/S1054-139X(13)00791-X/abstract?cc=y=

- An evaluator who is independent of grantees' organization to conduct the evaluation.
- Sufficient resources dedicated to the evaluation to ensure power and ability to maintain rigor. (OAH required grantees to allocate 20%–25% of their overall budget, but not more than $500,000 each year to the evaluation.)
- Placing a condition on a grant awards to stipulate that continued funding is contingent on meeting the standards set for evaluation rigor, and
- A commitment by the funder to hold grantees accountable to stated evaluation standards.
- Intensive programmatic and evaluation technical assistance for grantees to ensure high quality programs and rigorous evaluations.

Federally-Led Evaluations

There were two federally funded evaluation contracts managed by OAH. The first study, known as the *Evaluation of Adolescent Pregnancy Prevention Approaches (PPA)*, had two components: (1) an in-depth implementation analysis of six selected programs to examine the context and delivery of each program and provide a basis for interpreting estimates of program impacts; and (2) a rigorously designed impact study of each program using experimental designs and longitudinal survey data to assess the effectiveness of each selected program on its own, compared to a control group in the same site. In each of the evaluation sites, the PPA evaluation team worked closely with the local organization(s) implementing the programs to ensure successful execution of the study designs. In most sites, data was collected from the study sample through a baseline and two follow-up surveys. For two of the evaluations, three follow-up surveys were administered. The timing of the follow-up surveys varied across sites, ranging from 13 to 30 months after baseline, or from 12 to 24 months after program completion, in accordance with each program's theory of change and implementation schedule.

Another federally funded evaluation, known as the *Teen Pregnancy Prevention Replication Study Evaluation*, examined whether three evidence-based program models previously shown to be effective in a single study continue to demonstrate effectiveness when implemented with

fidelity across different settings and populations. Three replications of three evidence-based program models were examined for a total of nine sites. The study included a common set of outcome measures across all nine grantees participating in the TPP Replication Study, with minor changes to reflect differences in program focus. Data was collected at three points in time: when youth were enrolled in the study and before they were randomly assigned; between 6 and 12 months after random assignment (short-term outcome data); and between 18 and 24 months after random assignment (longer-term outcome data).

Building Capacity for Rigorous Evaluation – Challenges and Lessons from Cohort 1

Our experience with the first cohort of grantees of the OAH TPP Program brings to light several important issues regarding the capacity of the organizations to conduct rigorous evaluations. Many of these lessons have applications for government entities which are conducting evaluations in community settings. The challenges and lessons learned include the importance of providing for a planning and piloting period early in the grant cycle, measuring and monitoring fidelity and adaptations, ensuring high quality program implementation, incorporating evaluation effectively once program implementation has already begun, ensuring strong contrast between treatment and control, providing intensive evaluation technical assistance, and disseminating all evaluation results transparently.

Planning and Piloting Period Early

Grantees were not necessarily fully prepared for program implementation, so OAH required they engage in a planning, piloting, and readiness period for the first 6-12 months of funding. They needed sufficient time to conduct a thorough needs and resource assessment, hire and/or train staff appropriately and adequately, and make necessary revisions to ensure that the program materials were medically accurate. The pilot period also allowed staff to become comfortable with the program content, ensured the program was a good fit for the population, and identified any necessary adaptations to the program content or implementation. When grantees began full implementation of the program with large numbers of youth in their communities, they were better prepared to implement it with high fidelity and quality.

Measuring and Monitoring Fidelity and Adaptations

To increase the likelihood of replicating the same positive results associated with the original program evaluation, evidence-based programs need to be implemented with fidelity, i.e. the program is delivered in the way it was intended.

OAH defines implementing a program with fidelity as maintaining the core components of the program. OAH allowed grantees to make minor adaptations, if necessary, to ensure that the program was a good fit for the population being served as long as the adaptation did not compromise or delete the program's core components and was determined to be appropriate by the developer. Grantees were required to submit all adaptation requests, i.e. anything that was not implemented and evaluated in the original program model, to OAH for approval prior to implementation. In collaboration with the program developer, OAH staff determined whether the adaptation was appropriate and did not compromise the program's core components, before approval was granted.

OAH found that it was critical for grantees to set up a system for monitoring, analyzing, and using fidelity data to make continuous program improvements. OAH grantees were required to develop a fidelity monitoring plan, collect fidelity monitoring data, regularly review and analyze the data, provide feedback based on the data to implementation staff, use the data to see what was working well, and make continuous quality improvements.

Ensuring High Quality Program Implementation

It is not sufficient to make sure that programs are implemented with fidelity. In addition to fidelity, programs must be implemented with high quality to evaluate the true impact of the program. OAH required independent observations of program implementation to measure quality. Observers used a standard rating tool that assessed implementation on a series of constructs including participants' engagement, program facilitators' knowledge and enthusiasm, sessions' clarity and timeliness, etc. In addition, it is essential to ensure that participants receive the majority of the program. OAH found ensuring high dosage and

retention of participants, especially for programs that lasted several months or years, to be a significant challenge.

Incorporating Evaluation Once Program Implementation Has Already Begun

Recruiting grantees for the federally funded TPP Replication Study was complicated by the fact that most of their interventions were not designed to be implemented with the requirements of a rigorous experimental evaluation in mind. In some cases, the added burden of evaluation jeopardized the agreements schools or other partners had signed with grantees to implement the program (i.e. instead of a 100% chance of receiving the intervention, they could now potentially be randomized into the control condition; students would have to be surveyed; data would have to be provided by the schools). If the agreements could not be renegotiated, more time was needed to ensure sufficient sample sizes.

Ensuring Strong Contrast between Treatment and Control

Conducting a clean experiment is difficult when you have a very large federal grant program that spans 39 states and the District of Columbia, combined with other current or former federal funding streams for implementing pregnancy prevention programming. The control condition (i.e. the services that individuals would receive if they were not assigned to the program) for some of the evaluations was not sufficiently different from the program model tested to allow for a strong test of the model. In many cases, it took two years to achieve the necessary sample size for a rigorous experimental study in order to increase statistical power enough to detect impacts. In other evaluations, no change was able to be detected.

Providing Intensive Evaluation Technical Assistance

Grantee-led evaluations need continuous training and technical assistance to ensure their evaluations were designed, implemented, analyzed, and disseminated to meet the research quality standards of the HHS TPP Evidence Review. Grantees received intensive technical assistance on topics such as recruiting and retaining participants, coping with missing data, communicating with stakeholders, and calculating power or minimum detectable impacts.

Technical assistance was provided to all evaluation grantees through a contractor expert in random assignment evaluation, OAH's in-house evaluation specialist, and project officers. Monthly progress calls, OAH review and approval of all design and analysis plans, ad hoc calls as the grantees encountered real world challenges, and a vetting process of the final evaluation reports were given for each grantee evaluation.

Disseminating All Evaluation Results Transparently

No single program will work for everyone every time, and null or mixed findings, which are to be expected, are critical to continuing learning on which programs and approaches are most effective with different populations and in different settings. Nevertheless, there is a publication bias towards reporting the most positive outcomes. We need to be prepared for null results and to learn from the full body of evidence. To strengthen the science base, all of the results of the TPP Program evaluations have been disseminated and were communicated through individual evaluation reports, briefings, summary materials, and traditional and digital media.

Using Data and Evaluation to Improve Our Work

OAH invested in 41 rigorous evaluations during the first cohort of the TPP program from FY10-FY14. To facilitate learning in the field, OAH has taken numerous steps to ensure transparency and widely disseminate the evaluation results. The findings from 21 of the 41 rigorous

evaluations are featured in a September 2016 supplement of the *American Journal of Public Health*[4] and all the remaining evaluation findings are contained in reports held in a collection at the National Library of Medicine.[5] In addition, OAH developed several products (e.g. summary results tables, fact sheet, and infographics) and conducted numerous presentations to summarize and translate the findings. All of the products and individual reports are available on the OAH website at http://www.hhs.gov/ash/oah/oah-initiatives/tpp_program/cohorts-fy-2010-2014.html.

OAH used the evaluation results and lessons learned from the first five years of the TPP Program to redesign its funding strategy and evaluation framework for the second cohort as well as to assist grantees in selecting evidence-based programs that are the best fit for their community. The TPP Program is still a two-tiered program, with the majority of funding for replicating evidence-based programs and a smaller amount for developing and evaluating new and innovative approaches.

OAH found that supporting communities in implementing a holistic approach to replicating evidence-based programs in the highest need communities across the country can have a much larger, sustained impact. Instead of implementing a single evidence-based program at one time,

---

[4] *American Journal of Public Health,* Volume 106, Issue S1 (September 2016) is available at: http://ajph.aphapublications.org/toc/ajph/106/S1

[5] The National Library of Medicine may be accessed at: https://www.nlm.nih.gov/

youth need to receive evidence-based TPP programs at multiple times over the course of their adolescence. These programs should be delivered in an environment that is safe and supportive, trauma-informed and inclusive. Engaging community, families, and youth in the development, implementation, and evaluation of the program, and ensuring effective linkages and referrals to youth-friendly services in the community, is also important for increasing the impact of these programs.

Recognizing that not all communities are ready to implement evidence-based programs, especially in such a holistic way, OAH has set aside funding in Tier 1 to provide capacity building assistance (CBA) to those who are interested in, but not yet ready to do so. Eight grantees are each providing tailored CBA to at least three youth-serving organizations in their communities to implement, evaluate, and sustain evidence-based TPP programs. The CBA can focus on a wide range of topics, such as selecting and implementing evidence-based programs that are a good fit, making necessary adaptations that maintain fidelity and quality, and collecting and using performance measure data to make continuous quality improvements.

OAH has taken a similar approach with the Tier 2 funding for research and demonstration projects as well. Twenty-four grantees received funding to rigorously evaluate new and innovative approaches that are specifically designed to reduce existing disparities and fill gaps in the current evidence-base for teen pregnancy prevention, such as programs for vulnerable youth, including homeless youth, parenting youth, and those in juvenile detention and foster care. Another two organizations received funding to foster and support early innovations to prevent teen pregnancy, i.e. promising programs or technologies that need more support for development before they are ready for a rigorous evaluation. The intention is to bring all good ideas to a point where they can be evaluated in a rigorous way.

Continued Investment in Data and Evaluation

TPP Evaluation Framework FY2015-2019

While the structure of the OAH TPP Program has been further refined, OAH continues to use a multi-pronged approach to evaluation with the second cohort of new TPP grant projects. All grantees continue to be required to collect and report a uniform set of performance measures on reach, dosage, fidelity and quality, partnerships, training, and dissemination. Tier 2B grantees are also required to conduct a rigorous implementation and outcome evaluation. Evaluation grantees have been asked to archive their data, in a designated data archive, at the end of the grant period so that it is available for future researchers; it may also eventually be linked to other data sources. In addition, several federally-led evaluations are underway that are specifically designed to fill gaps in the current knowledge base. This "second generation" of federally-led research and evaluation activities address a more targeted set of research questions of significant practical relevance to OAH and the broader field. The five efforts to continue a cycle of investing, learning, and improving include:

- A study to identify and test (1) replications of commonly used but understudied evidence-based teen pregnancy prevention programs, (2) significant or meaningful adaptations to existing evidence-based approaches, and (3) selected core components, key activities, and implementation strategies of common programs using new data.
- A meta-analysis to extract the maximum value from previously funded evaluation efforts by synthesizing the evidence from across these evaluations. The study will examine what program or contextual elements make a difference to youth outcomes using reports and individual-level data collected from previous evaluations. This includes program approaches, characteristics, components, implementation factors, and participant characteristics associated with program impact.
- An intensive qualitative study to elucidate the feasibility and impact of scaling up evidence-based programs and the mechanisms through which 50 OAH grantees aim to reach the highest risk populations. This project will also collect data to identify a smaller

number of projects to implement rigorous evaluations, as well as design a rigorous, cross-grantee study, to assess the effectiveness of the larger holistic approach.

- A secondary data analysis of existing data from the OAH Teen Pregnancy Prevention Program to enhance understanding of interventions designed to reduce teen pregnancy and existing disparities on programs implemented with middle school youth. Funded by a cooperative agreement, the final report is expected in June 2018.
- An implementation and impact evaluation of a commonly implemented, but understudied teen pregnancy prevention program. In addition to collecting data to determine the effectiveness of this program in reducing births, pregnancies, STIs and risky sexual behaviors, the implementation study will examine whether using classroom teachers to implement the program, rather than outside health educators, is an effective mode of delivery. Final reports are expected in February 2021.

Critical Strategies for Building Capacity for High Quality, Rigorous Evaluation

Believing in a culture of learning, OAH is fortunate to have dedicated resources to support evaluation of approaches to teen pregnancy prevention. OAH is committed to rigorous evaluation and ensuring ongoing updates to the evidence-base, while also seeking to improve its own work. OAH has learned that investments are critical in the following areas:

1) uniform evidence standards and transparency;
2) evaluation training and technical assistance and capacity-building;
3) inclusion of a planning, piloting, and readiness period for all grantees;
4) program implementation with fidelity and quality;
5) collection of data to monitor program implementation; and
6) using data to make continuous quality improvements and expanding the evidence.

In the case of teen pregnancy prevention, it has been important to present a list of implementation-ready programs with evidence of effectiveness identified through a systematic and transparent review process. Communities need detailed information about the menu of

evidence-based programs, such as implementation, target population, and results, so that they can select programs that are the best fit for their community. There must be transparency in publishing all evaluation results that meet the standards for rigor, in recognition of the need to continue to develop the evidence base and learn from new evidence.

An established evidence review provides a standard, yet rigorous, set of expectations for future evaluations to meet if a program aims to be considered evidence-based. Clearly communicating evaluation expectations to grantees and providing technical assistance is essential to building their evaluation capacity. Grantees may not always have sufficient knowledge, skills, or resources to engage in high quality, rigorous evaluations. Yet, with an investment in continuous training and technical assistance, evaluations can be designed, implemented, analyzed, and disseminated to meet the research quality standards of the HHS Teen Pregnancy Prevention Evidence Review.

Similarly, communities need adequate support in preparation for implementation of their program. Grants need to include planning and piloting time so that effective implementation strategies and any issues can be identified early and the necessary adjustments made. Time to conduct thorough needs assessments, for example, will enable grantees to understand the needs of their population and make more informed decisions about which programs would be the best fit for their community and population.

Implementing programs with fidelity and quality, requires monitoring, support, and reliance on data for continuous quality improvement. To conduct evidence-based programs, funders need to commit to and ensure that communities have the capacity (staffing, training, use of data, etc.) to implement evidence-based programs with fidelity and quality, to minimize the adaptations made to the evidence-based program, and to ensure strategies for increasing retention of participants.

Finally, it is critical that we collect data on implementation of the programs and to use the data on a regular basis to make continuous quality improvements. This is not always easy, and may

require hard decisions about what to implement and how to implement. Nevertheless, attention to data allows us to monitor progress and better ensure we are meeting goals of improving the lives of adolescents across America.

In conclusion, evidence-based programming requires a significant investment in quality program implementation and a commitment to rigorous evaluation, which in turn means we need to build the capacity of community based organizations and use data to drive our decision-making.

Thank you for this opportunity to present. I am happy to answer any questions you may have.