Teen Pregnancy Prevention (TPP) Replication Study

Implementation Study Design Report

Contract No: HHSP23320095624WC
Order No. HHSP23337011T
May 9, 2013

Prepared for:

Lisa Trivits, Ph.D.
U.S. Department of Health and Human Services
200 Independence Ave., SW
Room 404E
Washington, DC 20201

Amy Farb, Ph.D.
Office of Adolescent Health
Office of the Assistant Secretary for Health
U.S. Department of Health and Human Services
1101 Wootton Parkway, Suite 700
Rockville, MD 20852

Submitted by:

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Table of Contents

Introduction

The Teen Pregnancy Prevention (TPP) Replication Study offers a unique and exciting opportunity to learn from the significant investment made in evidence-based teen pregnancy prevention programs through the HHS Teen Pregnancy Prevention Initiative. The goal of the evaluation—to contribute important information to the research base on teen pregnancy prevention programs—will be accomplished through a series of rigorous experimental design evaluations of a set of "evidence-based" programs that are being replicated by grantees funded by the Office of Adolescent Health (OAH). These studies will investigate whether evidence-based programs, when replicated with fidelity by grantees, produce behavioral impacts similar to those demonstrated in the original studies, and determine whether these impacts are sustained over a longer period than earlier studies have examined.

In addition, the design of the evaluation offers an opportunity to move beyond the question of the impact of a single replication of a program to look at variation in impacts for program models implemented in different contexts and/or with different populations. The impact study will be supplemented by a comprehensive implementation study that will allow us to examine the relationships between variation in impacts and program implementation. The implementation study will provide critical information about the contexts in which evidence-based programs are put in place, the challenges encountered, and the aspects of program implementation that are associated with program impacts.

The evaluation has two major components: a study of the impacts of replications of three program models, and an implementation study. This report focuses on our design for the implementation study; a companion report that will follow describes the impact study. The report begins with an overview of the evaluation as a whole. The chapters that follow present our plan for a comprehensive study of the implementation of three different program models, each replicated in three sites. We present the purposes and goals of the implementation study and the research questions that guide it. We propose a conceptual framework that draws on prior research to create linkages among the elements in the framework and that will be used to identify the data needed for the study. Data needs and sources, measures development and strategies for collecting the data are discussed, and a timeline for study activities is presented. Finally, we describe our approach to analyzing and reporting the data.

Overview of the Evaluation

The TPP Replication Study is a five-year evaluation of the effectiveness of replication of evidence-based programs designed to prevent sexual risk behavior and its consequences[1]. Through a series of rigorous experimental studies, the TPP Replication Study will test multiple replications of three widely-used evidence-based program models. The strategy of selecting multiple replications of each program model will allow for an examination of variation in impacts across replications of each program model and provide evidence about the generalizability of program effectiveness. A comprehensive implementation study will provide

[1] A detailed discussion of the rationale for the focus of the study on replication, and of the process of selecting program models and replication sites, can be found in a companion report that describes the design of the Impact Study.

information about the contexts in which evidence-based programs are implemented, the challenges faced in implementing them, and the aspects of program implementation that are associated with program impacts.

The program models that will be examined include: ***¡Cuidate!*** (an HIV/STD risk reduction program); ***Reducing the Risk*** (a sexuality education curriculum); and ***Safer Sex*** (a clinic-based HIV/STD prevention program for high-risk adolescent females).

The three programs reflect variation in program focus, service delivery strategy and populations targeted. ***Safer Sex*** (*SSI)* is a clinic-based program whose population target (female adolescents ages 14-19) is at maximal risk for teen pregnancy, since eligibility for services is confined to those youth who are sexually active at the time they seek services.[2] ***Reducing the Risk*** (*RtR*), by contrast, is a curriculum-based program, widely used in classroom settings (as well as some community-based settings) for students aged 13-19, a majority of whom may not be sexually active, even in high risk communities, such as those targeted by the TPP initiative. ***¡Cuidate!*** is a curriculum-based program for adolescents 13-19. The program is culturally tailored for Latino adolescents who are at high risk for HIV/AIDS, not all of whom are sexually active at the time they receive the program. Although it targets Latino youth, the program is typically delivered in English.

The three programs differ in their strategies for delivering service and the duration and intensity of the service provided. *SSI* provides one-on-one counseling, with minimal scripting of the encounters, to individual female youth in four sessions spread over six months; ¡*Cuidate's* six sessions can be delivered over two days or over one to six weeks, to small groups of 10-12 youth. The sessions are lightly scripted; topics for each session are identified and culturally-appropriate materials are provided. *RtR* has 16 highly-scripted sessions for groups that can range in size from 15 to 30 or larger. The program may be delivered over a semester or a shorter period of time, depending on the length of time allocated for the class.

Three replication sites have been selected for each of the three program models. Exhibit 1 summarizes the characteristics of the program models and their replications.

Program impacts will be estimated by using an experimental design in which youth are randomly assigned to treatment and control conditions. The unit of random assignment will be the individual or class, depending upon the program setting. Overall, the study will be based on a sample of approximately 9,000 youth in nine locations across the country. The evaluation will collect baseline information when youth are enrolled in the study and before the programs begin, short-term outcome data at a follow-up survey between 6 and 12 months post-baseline, and longer-term outcome data at a follow-up survey administered between 18 and 24 months post-baseline. Comparison of outcomes for program and control groups will provide important information about the effectiveness of the programs in reducing teen pregnancy and associated risk behaviors.

[2] The original *SSI* included females ages 14-23 hospitalized for treatment of an STD.

Exhibit 1. Key Features of Program Replications in the Evaluation, by Program Model and Replication Site

Program Model, Grantee	Program Description	Study Location	Target Population: Age	Target Population: Demographics (from proposal description)	Program Duration and Intensity	Program Setting	Program Delivered By
Reducing the Risk[3]	**Sexual health and risk prevention curriculum delivered to groups in schools or community settings**	**13 high schools throughout CA (46 classes)**	**High school students**	**62% white, 20% Hispanic, 9% Asian, 2% African American, 2% Native American**	**16 45-minute sessions, which can be doubled-up.**	**High schools**	**Teachers**
Better Family Life		St. Louis and East St. Louis, MO	9th graders	98% African American; low SES (75% eligible for free/reduced-price lunch in St. Louis City); high risk for teen births and STIs	16 sessions delivered over 8 to 16 weeks, depending on school schedule	Non-core classes in 6 high schools	Health educators trained and employed by BFL
LifeWorks		Austin, TX	9th graders (with small numbers of 10th and 11th graders)	75% minority youth, almost all below poverty level; teen pregnancy rates are increasing (37 pregnancies per 1,000 female high school students in '08-'09); high rate of STDs	16 sessions delivered over 8 weeks	Health classes in 4 high schools	Health educators trained and employed by Planned Parenthood (grant partner)

[3] Kirby, D., Barth, R. P., Leland, N., & Fetro, J. V. (1991). Reducing the risk: Impact of a new curriculum on sexual risk-taking. *Family Planning Perspectives, 23(*6), 253–263. This study found no effects after 6 months, but after 18 months, female, but not male, adolescents in the program who were sexually inexperienced at baseline were significantly less likely to report having had unprotected sex. No significant effects were found on sexual initiation, recent sexual activity, or pregnancy.

Program Model, Grantee	Program Description	Study Location	Target Population: Age	Target Population: Demographics (from proposal description)	Program Duration and Intensity	Program Setting	Program Delivered By
San Diego Youth Services		San Diego County, CA	9th graders (one school with 8th graders)	Very diverse; large Iranian population; youth at-risk for involvement with the juvenile justice system or mandated to receive services by a judge or probation officer; "teen pregnancy hotspots" identified by the state	16 sessions delivered over 8-16 weeks depending on school schedule	PE/health classes in 7 high schools	Health educators trained and employed by 5 agency grant partners
San Diego Youth Services		San Diego County, CA	Youth ages 13-19 enrolled in community agency programs (some diversion by juvenile justice system)	Very diverse, large Iranian population; youth at-risk for involvement with the juvenile justice system or mandated to receive services by a judge or probation officer; "teen pregnancy hotspots" identified by the state	16 sessions delivered over 2-3 weeks	5 community agencies	Health educators trained and employed by 5 agency grant partners
¡Cuídate![4]	**HIV/AIDs prevention program for small groups with emphasis on Latino cultural values**	**Saturday program serving neighborhoods in northeast Philadelphia**	**Adolescents 13-18 years of age, mixed gender**	**All Latino, 85% Puerto Rican**	**Six one-hour sessions that can be delivered over 2 days to six weeks**	**After-school programs or community-based organizations**	**Trained facilitators**

[4] Villarruel, A. M., Jemmott, J. B., & Jemmott, L. S. A randomized controlled trial testing an HIV prevention intervention for Latino youth. (2006). *Archives of Pediatrics & Adolescent Medicine, 160(*8), 772–777. This study found that adolescents in the program were significantly less likely to report having had sexual intercourse and multiple partners in the previous 3 months; they reported significantly fewer days of unprotected sex and more consistent condom use. No significant effects were found on condom use at last sex or the proportion of days of sexual intercourse that were condom protected.

Program Model, Grantee	Program Description	Study Location	Target Population: Age	Target Population: Demographics (from proposal description)	Program Duration and Intensity	Program Setting	Program Delivered By
Touchstone Behavioral Health	Approved adaptation to deliver in classes of 20-24 students with 2 facilitators	Phoenix, AZ	8th graders	61% Hispanic, 29% white, 7% African American; 18.5% below Federal poverty line	Approved adaptation added one session on pregnancy prevention. Seven sessions once a week for seven weeks	Non-core classes in 10 middle schools	Facilitators trained and hired by TBH
La Alianza Hispana		Boston, Chelsea and Lawrence, MA	9thgraders (some 10th and 11th graders)	62-78% Hispanic, 9-20% white, .4-25% African American; 68-88% free/reduced-price lunch	Six sessions once a week for six weeks	Non-core classes in 2 high schools, after school program in 2 high schools	Facilitators trained and hired by LAH
Community Action Program of San Luis Obispo		SLO county, CA	9th graders	29-47% Hispanic, 47-64% white, 1-3% African American; 35-50% free/reduced-price lunch	Approved adaptation added two sessions on STDs and pregnancy prevention. Eight sessions over 8 weeks	Pullout sessions during school day in 3 high schools	Facilitators trained and hired by CAPSLO
Safer Sex[5]	**HIV/AIDS prevention program for high-risk females ages 13-19**	**Urban children's hospital; adolescent clinic**	**Adolescent females who are not pregnant**	**49% African American, 18% Hispanic, 14% Non-Hispanic, White; all sought treatment for an STD at health clinic**	**Initial one-hour face-to-face session with three 30-minute booster sessions over six-month period**	**Health clinics**	**Female health educator**
Planned Parenthood of Greater Orlando		Orange County and adjacent counties, FL	Sexually active females ages 15-19, who are not pregnant	72% white, 21% African American, 25% Hispanic, 5% Asian; 41% of children living in economic hardship; high rates of STDs		Two PPGO reproductive health clinics in Orlando	Health educators trained and hired by PPGO

[5] Shrier L.A., Ancheta R., Goodman E., Chiou V.M., Lyden M.R., & Emans S.J. (2001). Randomized controlled trial of a safer sex intervention for high-risk adolescent girls. *Archives of Pediatrics & Adolescent Medicine, 155(*1), 73-9. This study found no effects one month after the program, but six months after the program, adolescents who participated in the program were significantly less like to report having had another sexual partner, aside from their main partner, in the prior six months.

Program Model, Grantee	Program Description	Study Location	Target Population		Program Duration and Intensity	Program Setting	Program Delivered By
			Age	Demographics (from proposal description)			
Knox County Health Department		Knox County and adjacent counties, TN	Sexually-active females ages 14-19 who are not pregnant	89% white, 9% black, 19% females 15-19 are Latina; poverty rates up to 34% for children under 18; many teens from high risk situations; serve children in state custody		16 reproductive health, adolescent health clinics	Health educators trained and hired by Knox County Health Department and grant partners
Hennepin County Health Department		Hennepin County, MN	Sexually-active females ages 14-19 who are not pregnant	32% African American, 10% Latino, 46% Caucasian; large disparities in family income by race/ethnicity; sites selected for program implementation have teen birth rates approaching or exceeding the national teen birth rate		20 reproductive health, adolescent health, school-based health clinics	Health educators trained and hired by Hennepin County and grant partners

The Implementation Study

OAH's Teen Pregnancy Prevention Program places a heavy emphasis on the replication of evidence-based program models, also requiring that a substantial proportion of the grantees that replicate such programs conduct rigorous evaluations of their effectiveness. In addition, the Replication Study proposed here will conduct randomized experimental evaluations of nine programs funded under the OAH TPP program, all but two of which would not otherwise have been the subject of a rigorous evaluation. These efforts, in addition to the other federally-funded evaluations associated with the initiative, will provide a vast amount of valuable information to the field. However, as Fixsen and Blase (2008) point out, more and better research on a program does not, by itself, lead to successful replication on a larger scale or help translate "science to service".

A comprehensive implementation study addresses this issue first by providing an in-depth description of a program as it operates in a real-world setting, identifying the challenges faced, the barriers encountered, and the strategies employed to address them, and secondly by assessing or evaluating the extent to which the program met its own goals and the extent to which aspects of the program itself, its sponsors or the environment in which it operates help or hinder its full implementation. Finally, such a study, if carefully designed to do so, can be explanatory, that is it can help illuminate factors that affect the outcomes the program is seeking to achieve.

The goals of the Implementation Study are:

- To provide an in-depth description of the intervention as planned and implemented in each of the replication sites for the three models;
- To document the extent to which program models are implemented with fidelity and are able to meet their performance goals;
- To examine barriers and challenges to implementation in each of the sites in order to arrive at a qualitative understanding of why replication efforts did or did not reproduce the impacts reported in the original study;
- To identify and describe the services available to and used by youth in the control groups; and
- To link aspects of program implementation to variation in program impacts, in the event that the impact study identifies such variation.

Underlying these goals are two questions that are shared by many research fields, namely: the extent to which multiple high-quality replications of a program model are feasible; and the factors both internal and external to the program that affect replication.

Research Questions for the Implementation Study

The program models and the replication sites selected differ in the scope of the intervention, the populations they serve, and, in some cases, the setting in which the intervention is delivered. For these and other reasons, the implementation study for each site will have some unique features. Notwithstanding, an overarching set of questions drives the design of the implementation study, the types of data needed to answer the questions, the strategy and measures used to collect the data, and our approach to analyzing the data.

The first three questions address the issue of feasibility of high-quality replication. They are as follows:

1. ***To what extent was the program model in each site implemented as planned?***

   1a. To what extent did implementation vary across different locations within a site and across sites implementing the same model?

   1b. What aspects of implementation varied within and across sites?

   1c. What were the reasons for the variation?

Answering this question requires, first, a clear understanding of the goals and design of the program model being replicated: the rationale for the approach, underlying theory (including theory of change), target population, essential components, staff and other resource requirements, hypothesized outcomes. The next step in the process is to develop an understanding of the planned replication in each site. Although OAH mandated fidelity to the model, this does not change the fact that the planned replication may have additional outcome goals (reduction in teen pregnancy, for example, vs. reduction in STIs only), target a different population, choose one or more settings that differ from the original, and make other approved adaptations to the original model. Understanding these differences, where they exist, will allow us to delineate exactly what was being proposed in each replication site and then to examine how actual implementation differed from the plan for each site.

The investigation of change and adaptation, which begins with the replication plan outlined in the proposal, will move through two major subsequent stages: adaptations and changes made after the pilot year, and before full implementation; and subsequent changes made during full implementation.

Although the first question correctly assumes that we can arrive at an estimate of how closely the program implemented in each replication site matched what was planned, it is plausible that, even within a site, implementation might vary by location. In addition to looking at the *average* level of implementation, it will be important to try to capture variation in implementation within a site, for those programs that are implemented in two or more locations, sometimes by one or more partners. Variation in implementation within a site could occur at the partner agency level, at the individual health educator level or at the school, class or group level. Examples of such variation have already presented themselves: a program implemented after school hours in one location has considerably more difficulty retaining participants than in other locations where it is delivered during the school day; health educators encounter resistance and barriers to implementation in one school but not in others. Areas in which variation might affect implementation include: the local context; levels of experience with youth programs; administrative structures; staff qualifications and retention; barriers for potential recipients of the intervention etc.

Our ability to investigate within-site variation is largely dependent on the extent to which the fidelity and performance data collected and reported to OAH reflect such variation. Interviews with program staff at all levels will explore the topic of variation in implementation and the reasons for it, but would ideally supplement the data collected systematically over time, rather than replacing it.

The replications of evidence-based program models were not intended to be medical replications, (i.e., replicating the "treatment" with an identical population in identical settings), but rather to explore the use of the interventions with different populations and/or a wider range of settings. Nevertheless, all the grantees were charged with maintaining fidelity to a set of core elements defined by OAH in consultation with the program developer. The next set of questions investigates the extent to which this effort succeeded.

2. ***Was the program model implemented with fidelity?***

   2a. Was the level of fidelity uniform across different locations within a site?

   2b. How and to what extent were measures of fidelity used to provide feedback and retraining to staff?

   2c. How long did it take to achieve a satisfactory level of fidelity?

Fidelity of implementation is the extent to which delivery of an intervention adheres to the protocol or program model originally developed. OAH has paid careful attention to the issue, providing monitoring, guidance and tools for maintaining fidelity, as well as an array of fidelity measures and a schedule for reporting them back to OAH. What the evaluation can do with the same data is to look at variation within a site, at how local evaluators and/or program staff chose to use fidelity data to improve program implementation and at how the feedback process affected implementation fidelity over time. Later in this document, we discuss the issue of whether measures of fidelity can be used to explain variation in program outcomes.

While the first two sets of questions will provide information that will allow us to describe in detail, for each program model, variation in implementation across and within replication sites, the next questions examine the challenges or barriers that programs faced in implementing their plans, and the solutions they devised, when possible. These challenges may provide explanations for the adaptations made to the plan and will allow us to investigate plausible reasons for variations in the strength of implementation.

3. ***What challenges or barriers to implementation did grantees encounter?***

   3a. How did they deal with challenges and/or barriers?

   3b. What changes/adaptations were made to the original plan?

   3c. What was the rationale for the changes?

   3d. Did the changes strengthen or weaken program implementation?

Grantees may experience challenges or barriers to implementing the program model across the site or in specific locations, and may make changes or adaptations to the original plan in order, for example, to recruit and/or retain participants. Understanding program responses to these challenges will help us to refine the definition of what was tested and address real-world questions about what it takes to replicate a program model that may have been developed many years earlier, for somewhat different populations and in a different context. Although many of the changes will be intended to strengthen implementation of the program, it is important to

recognize that some changes that are made because of circumstances beyond the program's control may adversely affect implementation of the model as planned.

The remaining two research questions address the issue of the factors that potentially mediate program outcomes.

4. ***What aspects of the program, as implemented in a site, or of factors exogenous to the program, seem to be associated with greater impacts (or with lack of impact)?***

   4a. What were the relevant experiences of members of the control group?

Fidelity to a program model is not the only aspect of implementation that might affect participant outcomes. The conceptual framework shown in Exhibit 2 will guide the exploration of factors that might, alone or in combination, provide possible explanations for variation in program impact. ***Readiness***, both on the part of the grantee and partners and the program model itself, the ***context*** in which the program is implemented, and the extent to which ***supervisory staff*** monitor and support staff who deliver the program may all affect the ***fidelity and quality*** of program implementation of the program and force ***adaptations*** that strengthen or weaken the program. In turn, the strength and quality of program implementation influences its ability to attract and retain participants and their ***responsiveness*** to the program's messages – critical antecedents of program impact.

One external factor over which the program has no control is the receipt by members of the control group of services similar to those offered by the program, which may reduce or eliminate program impact. Agency and school staff can provide information about other programs or services that are being provided to or are available to control group members. For information on services that individual youth receive, we will rely on data collected for the impact study from surveys of youth in both the control and treatment groups. None of these data are very fine-grained, but for the purposes of the study, our main concern will be to identify ***systematic*** provision of comparable services. Both treatment and control group members may be exposed to multiple sources of relevant information or even receive similar services to those offered by the program, but these will only affect our ability to detect program impact if there is systematic and differential use of them by the control group members.

Although we cannot attribute causality to any of the factors that seem to be associated with differential impacts, we can investigate them and provide qualitative evidence to support one or more hypotheses.

The final question for the Implementation Study reflects interest in a more systematic linkage of variation in program impacts to variation in program implementation.

5. ***How does variation in aspects of implementation influence the impact (or lack of impact) of interventions to prevent teen pregnancy?***

Unlike the prior questions, which focus exclusively on the group that received the intervention (or the one question that deals with the experience of the control group), answering this question will require the use of data on both treatment and control groups (collected in the surveys conducted for the Impact Study), capitalizing on the power of the experiments, and using recently-developed analytic techniques. Although earlier research has tried with some success to

tie key aspects of implementation such as fidelity to a model to program ***outcomes,*** only in the last few years have there been efforts to link variation in implementation to program i***mpacts,*** in order to provide strong evidence for policymakers and program operators on those aspects of program models that can be identified as key for effective intervention. The conceptual framework described in the next section builds on and cites the work of researchers who have linked elements of program implementation to program outcomes. Some of the newer research linking implementation to impacts is cited in a later section that describes a potential approach to this analysis.

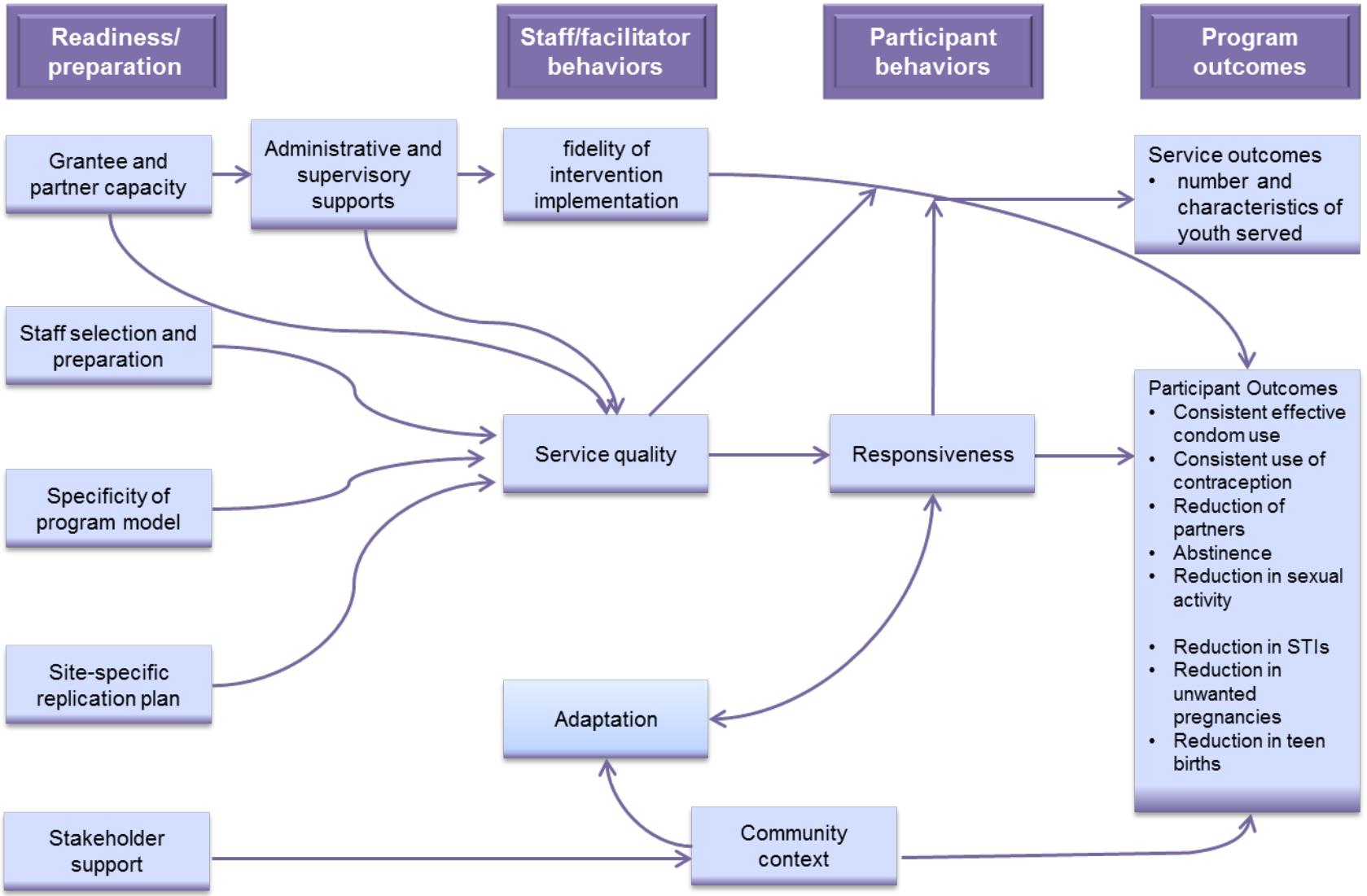
Conceptual Framework for the Implementation Study

To guide the specification of data needed to address the research questions, we propose a framework that builds on the work of Berkel and her colleagues (Berkel et al., 2011) and others, to identify aspects of implementation that have been shown to affect program outcomes, as well as the factors internal and external to the grantee that affect implementation. Exhibit 2 presents the proposed framework.

Moving from the right-hand side of the diagram, the participant outcomes shown are, first, the major outcomes specified by OAH. In addition, a set of behavioral outcomes are necessary precursors of reductions in teen pregnancies and births and in STDs. They are: abstention from sexual intercourse, or reduction in sexual risk behaviors.

Berkel and her colleagues propose four behavioral mediators that reviews of prior research (e.g., Dane and Schneider, 1998; Durlak and DuPre, 2008; Dusenberry et al., 2003) have linked to participant outcomes. Three are staff behaviors, one is participant behavior. The three staff behaviors are: fidelity to the program model; quality of services; and adaptation. The participant mediating behavior is responsiveness. ***Fidelity of implementation*** is the extent to which key program components are delivered as prescribed by the program developer, in terms of content, delivery methods and the amount of time spent on each component. ***Quality of service*** refers to the instructional approach and the skill with which facilitators or educators deliver program material and interact with participants. ***Adaptation*** refers to changes made to the program as planned, such as, for example, changing recruitment and retention strategies, adding materials that are relevant to participants' lives or that fill a gap in the existing curriculum. In TPP, these adaptations must be approved by OAH and would be seen as enhancing the evidence-based curriculum. (Of course, adaptation can have negative consequences if the changes are a result of external pressures or inadequate staff training, and the study must take this possibility into account in developing measures.) Berkel and her colleagues make a convincing argument that the quality of the services delivered and adaptations that enhance the strength of the intervention produce positive response on the part of participants. ***Responsiveness*** includes: attendance at program sessions; active participation and engagement in program activities; and satisfaction with the program. Participant responsiveness and quality of service interact with fidelity to produce the desired outcomes, in terms of both service outcomes – the number and characteristics of youth served, and participant outcomes – reduction in teen pregnancy, teen births and STDs.

Exhibit 2: Implementation Framework



Most recent implementation research identifies the front-line staff who deliver the program as the "drivers" of implementation. The actions and processes that program administrators put in place to support the work of these front-line staff is crucial to successful implementation. The ***administrative and supervisory processes*** that foster positive staff behaviors include: decision-making and problem solving processes that include front-line staff; clear rules and performance standards; in-service training, consultation and coaching that is responsive to staff needs; fidelity and performance monitoring; regular feedback to improve performance; and effective work with external systems and agencies to ensure needed support for the program.

Although readiness and preparation are not always part of the discussion of implementation, we believe that they are crucial to the ability of grantees to implement the program as planned. Indeed the requirements of the grant application and the provision of a planning and pilot year for all grantees make it clear that OAH also perceived the importance of these precursors of implementation. We have identified five aspects of ***readiness and preparation***: 1) the capacity of the grantee and any partners; 2) the qualifications, experience and preparation of staff selected to deliver the program; 3) stakeholder awareness of and support for the program; 4) the specificity of the program model; and 5) the site-specific plan for its replication. Grantee and partner capacity includes: infrastructure, experience in delivering similar programs or serving youth; financial and programmatic resources; policies that support or might undermine the program; and the agencies' standing in their community. The leadership provided by the grantee's senior staff and their vision for the program, understanding of community needs and ability to select a program model that meets those needs, recognition of and preparation for potential challenges and their ability to identify and involve key stakeholders also reflect grantee capacity. Staff capacity includes: the process and criteria used to recruit and select front-line staff who will deliver the program or to select candidates from existing staff; their training; and the extent to which they share the program's goals and are committed to its implementation. Stakeholder support includes understanding the risks and benefits of the program and the existence of one or more program champions.

Successful implementation is also affected by the specificity of the program model selected , i.e., the extent to which its theoretical basis is clear, goals, core components, curriculum content and staff requirements are clearly set forth, training is available and effective, and the program has clear and effective written manuals and other guidance. All of these elements determine whether there exist clear guideposts for replication. The site-specific plan for program replication, which includes approved adaptations, is the standard for judging whether the program is successfully implemented. A realistic plan for staffing, recruiting participants and choosing a setting for the program, among other elements, is more likely to be successfully implemented.

A factor that affects the ability to fully implement a program and may also directly affect outcomes is the external context in which the program is operating, whether this is a community, one or more neighborhoods, or a school district. Although most of the replications are by design serving disadvantaged and high-risk youth, even areas of high poverty differ in levels of social cohesion, the availability of resources for youth, especially sexual health services, the level of support for or opposition to the program, the presence of active advocates for the program, and the behavioral and social norms that operate in the community. Understanding this context will be important for our understanding of the factors that support or impede implementation.

Finally, and because the implementation study is taking place in the context of an impact evaluation of the replications, it includes an element that does not affect implementation but may affect our ability to detect outcomes. The services that youth in the control group receive, the counterfactual, are documented as part of the implementation study, not because they relate directly to it, but because they are essential for the assessment of program impact.

Data Sources

To obtain information on all these topics we will tap into a wide range of sources. These include: site and program documents, such as the proposal and revisions to it, semi- annual reports and refunding applications, internal reports and records; site documents, reports and records; program officer reports; notes from site investigation, selection and negotiation; discussions with program staff and stakeholders during site visits: on-site observations; program performance and fidelity data collected by the program or the local evaluator; staff and youth focus groups; and the youth surveys developed and used for the impact study. Exhibit 3 links data needs and sources.

Exhibit 3: Implementation Framework Elements, Constructs and Data Elements

Readiness/Preparation	Implementation of Intervention		Community Context	Participant Behavior
Construct: *grantee and partner capacity* **Data elements** • grantee and partners' position and role in the community • grantee and partner's prior program experience • size and resources of grantee agency • method and appropriateness of program selection	**Construct: *grantee and partners' administrative and supervisory processes*** **Data elements** • Amount and type of in-service training for staff • Amount and type of consultation and coaching • Amount and type of performance monitoring, staff evaluation and feedback	**Construct: *staff attitudes/satisfaction*** **Data elements** • Staff workload • Staff commitment • Staff satisfaction with program model • Perceived adequacy of training and preparation • Perceived adequacy of monitoring and feedback	**Construct: *level of community risk*** **Data elements** • Community demographics • Youth population demographics • Community needs • Community cohesiveness vs disorganization • Community values and norms	**Construct: *participant responsiveness*** **Data elements** • attendance • engagement • satisfaction with program
Construct: *staff selection and preparation* **Data elements** • Strategy for recruitment or selection of staff • Qualifications of program staff selected • Initial staff training • Proposed staffing structure and workload	**Construct: *extent to which intervention was implemented as planned*** **Data elements** • Program implemented in planned settings • Changes in program settings and rationale • Intervention staffed as planned • Changes in staffing and rationale • Staff turnover and reasons • Effect of staff turnover on program implementation • Characteristics of population served vs. planned • Barriers to recruitment and revised recruitment strategies • Problems with participant retention and strategies to overcome them	**Construct: *fidelity of implementation of the intervention*** **Data elements** • Average and range of fidelity scores for individual staff • Average and range of fidelity scores by session • Areas in which achieving fidelity is difficult • Use of fidelity measures to provide guide staff • Perceived adequacy of measures • Staff reaction to fidelity measures • Perceived effect of improvement in fidelity on participant engagement/retention	**Construct: *community resources*** **Data elements** • Availability of programs for youth • Availability of reproductive and other health services for youth • Barriers to accessing services	

Readiness/Preparation	Implementation of Intervention		Community Context	Participant Behavior
	• Schedule for program activities (number of sessions, length, duration) • Challenges presented by schedule and effect on program implementation • Program components and activities implemented • Modifications in activities and rationale • Gaps in/problems with program content/length/number of sessions	**Construct: *Adaptation of program model*** **Data elements** • Number and nature of adaptations requested and approved • Rationale for approved adaptations • Number and nature of adaptations requested but denied • Rationale for denied adaptations. **Construct: *quality of services*** **Data elements** • average and range of observation scores for individual staff • average and range of observation scores by session • Schedule for observations • Perceived variation across staff • Use of observation measures to provide staff guidance • Perceived adequacy of observation measures • Staff reaction to observation measures • Perceived effect of improvement in quality of service on participant engagement/retention		

Readiness/Preparation	Implementation of Intervention		Community Context	Participant Behavior
Construct: *specificity of/support for program model* **Data elements** • Adequacy of materials for program and reference • Adequacy of other support for implementation				
Construct: *site-specific replication plan* **Data elements** • Proposed replication plan with approved adaptations to model • Post-pilot replication plan with approved changes/adaptations and rationale				
Construct: *stakeholder support* **Data elements** • support for grantee from schools other organizations • opposition to grantee and source of opposition				

Measures

The primary measures developed for the implementation study are topic guides for telephone and in-person interviews and focus group guides for on-site data collection. A set of generic guides were developed, submitted and approved by OMB. It will be important to customize the guides, first for each of the three program models which, as shown earlier, differ from each other in several important aspects, by adding probes, defining discussion topics, changing terms and language used to reflect and capture unique aspects of the program model. A second level of adaptation may be necessary to reflect and capture variation at the individual site level. For example, if a program like *SSI* is implemented in different types of clinics within a replication site, implementation may pose different challenges from those encountered when the program is implemented in two clinics operated by the same sponsor.

Fidelity and ***quality of service*** are critical aspects of implementation. In the current study, fidelity and quality measurement is potentially facilitated by the OAH requirement that grantees use OAH provided fidelity measures that include some assessment of quality of service, as well as performance measures that, among other things, document recruitment and retention of program participants. The use and reporting of these measures on a required schedule is mandated by OAH. They may be administered by local evaluators or administrative staff (though OAH guidance recommends that observations be administered by an external and objective individual who has been trained in their use). Our plan for obtaining these data differs by site and is intended to place the minimum burden on program staff. Where feasible, we will obtain data directly from each program at three six-month reporting points (May 2013, November 2013, May 2014) with a possible fourth reporting point (November 2014) for the three Safer Sex sites, where recruitment into the study may continue longer.

We need to obtain data directly from the programs because, for the most part, grantees that use electronic record-keeping aggregate the data before uploading it to the RTI website. For the implementation study, we need disaggregated data that we can then aggregate, by school or health educator, for example. To meet the needs of the impact study, we need data on attendance that can be linked to individual students in the study. Finally, we need data only on those youth and schools that are part of the study, usually a subset of those served by the program.

The Safer Sex sites present the fewest challenges since the fidelity and performance data are entered directly into the electronic participant tracking system and we simply need permission to access and download them. Other sites that enter data into their own electronic system will also pose few, if any problems. The challenge will arise in sites that use paper records (fidelity checklists, attendance logs) and then have agency staff enter the data into the RTI system. If grantees have entered disaggregated data directly into the RTI website, we will determine the best way to access those data.

As we noted earlier, the data collected by grantees and reported to OAH is potentially valuable for our investigation of this topic. However, because analysis of the fidelity data will be both expensive and time-consuming, we will conduct some preliminary analysis of data reported for one of the three replications of each model, where we have prior knowledge of variation in implementation, to determine the extent to which the data accurately reflect the variation.

In addition, we have developed an interview guide for use with program directors that explores the issues surrounding efforts to replicate with fidelity and to balance those efforts with appropriate adaptations that meet the needs of the populations they serve.

The data from the systematic observations conducted by supervisory staff or local evaluators will be very useful in any assessment of the quality of service. In addition, we plan to conduct observations of program activities during site visits wherever possible. While these observations will necessarily be limited to a single session in any one site (though it may be possible to observe the same session in more than one school), they will provide some independent validation of the observations conducted by grantee staff. These observations will be easiest to schedule and conduct in Reducing the Risk classrooms, must be approached more carefully in ¡Cuídate! settings (when the program is implemented in small groups) and not possible for Safer Sex. *SSI* is offered in one-on-one counseling sessions where an outside observer would be maximally intrusive and almost certain to change the nature and effectiveness of the interaction. The grantees themselves most often do not conduct observations for those reasons. If it is determined that observations are feasible in most settings, we will plan to use the same observation measures that the programs are currently using.

We have developed focus group guides for use with program participants, as well as with program staff. In the case of program staff, such group discussions might be an effective and efficient way of capturing their perspectives in larger replication sites. The alternative approach would be to sample staff in programs with large numbers of front-line staff, and conduct individual interviews. If focus group discussions with program staff and/or participants are determined to be a potentially useful source of information, we will a refine the measures submitted to OMB, adding program- or model-specific probes where necessary.

The large amounts of information that will be abstracted from existing documents need to be recorded in a systematic fashion. A template for recording information from multiple sources, including grantee reports and other documents, has been developed.

Data Collection

A substantial portion of the data needed for the implementation study will be extant data (funded grant proposals, reports, records, program officer notes and reports, information from the recruitment and selection process, fidelity and performance data, youth surveys). In some cases, the data will be obtained from each grantee. We will also submit a request to OAH, for copies of site reports, with an explanation of how the information will be used and the research questions it will help us address. A similar request will be made for copies of applications for continuation funding and responses to questions about them.

In addition, in-person visits to and telephone contacts with grantees will provide essential information about the process of program implementation at each replication site. Initially, interviews with agency and program staff will be conducted by telephone, beginning in November 2012 and completed by early February 2013.

A round of visits to each site will be conducted in the fall of 2013 (except for two sites where the study sample will be recruited by Spring 2013. Those site visits need to take place in the spring of 2013). We will work with grantees to time the visit so that it is maximally useful (i.e., the

program is being offered, in the case of Reducing the Risk) and develop a schedule for the visit that places the least burden on staff while allowing us to collect the necessary information.

Staffing the Data Collection

The telephone interviews will be conducted by a small group of study staff, led by the Abt Task Leader and the Abt Project Director. Each of the replication sites has been in weekly or biweekly contact with a senior study staff member (the Abt site liaison), and those staff, in most cases, will take the lead in conducting interviews with senior grantee and partner staff. In each case, the site liaison will be supported by a second member of the study staff, who will sit in on some of the early interviews and may conduct subsequent interviews with frontline staff.

To ensure that respondents feel free to respond openly to our questions, we will restrict the number of staff present at each interview to two, usually the members of the team assigned to a specific replication site. However, we also recognize that not all staff are equally skilled at eliciting the kind of information we are seeking. For this reason, the Task Leader or another senior staff member will sit in on the initial interviews conducted by each member of the data collection team, to monitor the interviewer's establishment of rapport with respondents, use of probes and follow-up questions, and skill in eliciting full and responsive answers to the questions posed. Feedback on these issues will be provided immediately after the interview and before any other interviews are conducted.

The site visits next year will be conducted by the same two-person teams, whenever possible. The teams will visit each replication site for up to three days (site visits might be shorter in the one or two sites that are implementing the program in two or fewer locations). Our goal throughout will be to ensure that both members of the site visit team brings to the effort a detailed understanding of both the program model and the replication site.

Preparing for Data Collection (telephone interviews and site visits)

In preparation for the interviews, we will prepare a site-specific logic model that builds on the program logic model. This complements the framework shown in Exhibit 2, which is intended for use across all program models and sites. Both will help study staff understand the intended uses of the information they are collecting. In addition, using the funded grant application and any other documents available at the time, we will prepare preliminary program profiles (i.e., a description of the intervention as originally planned) and a description of the control condition, as we understand it from our discussions and negotiations with the grantee. These descriptions will provide a starting point for the discussions with program staff about changes to the plan and the reasons for them.

To the profiles will be added customized discussion guides and protocols tailored to each program model, replication site and group within the site. For the site visits, focus group guides, similarly tailored to the program models and the participants (youth and/or health educators/facilitators) will be prepared. Instructions for the use of interview measures will be included in a ***Data Collection Guide,*** for the telephone interviews that will be conducted this fall. The guide will provide instructions on scheduling and conducting interviews, and on the format in which interview notes should be summarized. An expanded version will be incorporated into a ***Site Visit Guide*** for the site visits next year. This guide will detail the steps to be taken in order to set up site visits, conduct interviews, structured observations and group discussions. The

Guide will also be used to train study staff in how to conduct data collection activities and record information effectively and consistently.

The teams identified for the telephone interviews this fall will participate in a one-day training session before any interviews are scheduled.

Scheduling Visits

Following procedures outlined in the ***Site Visit Guide***, Abt Associates staff will make arrangements with grantees to conduct the site visits. One member of each team will be assigned to coordinate the visit, working with the individual whom grantees have identified as the key contact. Abt staff and the grantee liaisons will discuss the items that need to be included in the site-visit agenda, but we will rely a good deal on their judgment in planning how to schedule the individual and group discussions, focus groups and observations we want to conduct.

Team members will be given checklists to ensure that each necessary step has been taken to prepare for the site visit (including the process for obtaining parental consent for youth to participate in focus group discussions, any advance steps such as finger-printing of site visitors that schools may require, preparation of a site visit schedule to be shared with grantee and other staff, confirmation of dates and times of meetings with individuals and groups). Schedules and plans will be shared with federal staff in a timely manner to allow them to participate in the site visits if they wish to do so.

One week before the first round of site visits, we will conduct a two-day training. Since team members will be experienced site visitors, we expect these sessions to be highly interactive for the most part, after an initial briefing on the purposes of the study and the research questions it will address. Training will include drawing attention to key aspects of implementation to which site visitors should be especially attuned and an overview of the purpose and intended use of the various types of data. Training in observation protocols (to the extent that observations are feasible) will include examples from previous site visits involving similar program models. In addition, site visits will be conducted by pairs of observers, who will conduct observations together and discuss their coding. The entire team will meet as a group to clarify coding after each pair has observed at least one session. In addition, the observation protocols provided by OAH are well-specified and lend themselves to high inter-rater reliability.

Documenting the Interviews and Visits

As we noted earlier, the information gathered through telephone interviews and site visits, though of critical importance to the study, is only a part of the data that will be collected; an equal amount of information will come from fidelity and performance data and from program documents. For this reason, interviewers and site visit staff will not be asked to transform their interview and focus group notes, and observation data, into a conventional site visit report. The information they collect is for our purposes raw data to be entered, with data from other sources, into a database by trained coders. We will therefore provide each interviewer and site visitor with a structured framework within which they can place their interview notes in preparation for coding.

Coding Implementation Data

The next important step will be to institute a systematic and comprehensive approach to organize and analyze the quantitative and qualitative implementation data. This includes developing a

well-defined codebook for use with a qualitative software package, and training coders to identify content with a high level of inter-rater reliability. The codebook will describe the information that coders will "tag" for abstraction across all documents included in the study. Structured to capture all of the important implementation data available, codes will identify both quantitative and qualitative data. Areas of initial coding will include: characteristics of the three pregnancy prevention programs implemented; key characteristics of participants; and themes related to the research questions (e.g., factors challenging implementation, factors facilitating implementation). In addition to coding for categories that are defined prior to reviewing the data, we expect to also use an inductive coding process to include new themes that emerge from the data.

Trained coders will apply this codebook to all implementation reports and relevant documents using NVivo, a software package that facilitates the search and retrieval of content by code. Especially efficient when used with a large quantity of data, this qualitative software package provides an essential tool for data management and analysis. The codes can be refined over time, reorganized in new ways, subdivided or merged, in NVivo; this enables us to examine relationships between and among implementation success and various key characteristics about the program, its participants, communities, etc., to help us understand the important differences in implementation. In addition, NVivo facilitates the "quantification" of qualitative data by rapidly producing code counts and matrices reporting the frequencies with which codes overlap, and may be imported into quantitative software packages for statistical analysis. All coders will receive training so that they use the codebooks reliably and interpret the data in a similar manner. Where possible, all materials will be double coded and inter-coder reliability queries will be used to ensure consistency across coders and the content abstracted.

We have used NVivo to illustrate the process we plan to use because it seems to be the best suited for the purpose of this study. However, we are in the process of reviewing with Abt experts on this system the types and volume of data that we will collect, the ways in which the system will allow us to manipulate the data, and any caveats they may have that would suggest reconsideration of this decision.

Analysis and Reporting of Implementation Data

Our analytic strategy moves from site-level descriptive analyses through qualitative evaluative analyses at both the individual site level and across replication sites within a program model, to quantitative analyses that use outcome data on all study participants (both treatment and control) to link variation in implementation to impacts on outcomes for youth. Below, we describe our approach to each type of analysis.

Descriptive Analyses

The first step in the implementation analysis is to construct a site-specific description that "tells the story" of what happened in a comprehensive way, tracing the process of replication and the context in which it occurred. The description has two main components: one non-quantitative (the program narrative) and one quantitative (descriptive statistics). The framework proposed in Exhibit 2, as well as the program's theory of change, will frame the analysis. The analytic meetings held after each round of site visits will help to identify important topics or themes that emerged across replication sites. From the frameworks and the analytic meeting, a set of detailed research questions will be developed.

The software we propose to use facilitates combining information from different sources about the same topic. No one informant will provide comprehensive descriptions of the entire program and its results. The story of the program, from planning and preparation through start-up to full operation and outcomes, will need to be built up from multiple partial views. The account will include a discussion of the challenges encountered by program staff and the strategies they developed to address them. When accounts of the same topic agree, a simple summary of the topic can be developed. Where accounts of the same topic disagree, the analyst must decide on the meaning of the disagreement and document it.

The narrative will be supplemented by diagrams and timelines, as well as tables that summarize topics such as participant characteristics, participation in the program, fidelity to program components, changes in levels of fidelity over time. The counterfactual condition in each site will also be described.

Evaluative and Explanatory Analyses

For the next set of analyses, we will look both within a site and across the three replications of each program model to assess the extent to which the program was implemented as planned and to identify potential explanations for variations in implementation and in participant outcomes. The standards by which the adequacy of implementation is judged include: the requirements of the program model; the grantee's own plan for replicating the model and theory of change; OAH expectations for fidelity and performance; and stakeholder opinions and judgments.

Understanding why a replication is not working as planned is a particularly useful function of implementation research since it allows both policymakers and program operators to make needed adjustments either to the model itself or to plans for future replications. We will examine the ways in which differing levels of grantee and partner readiness and preparation, the appropriateness and adequacy of the program model selected and the complexity (i.e., number of different locations within a site) of the plan for implementing it, whether or not it is realistic, as well as external factors such as community norms and the availability of sexual health services in the community affected fidelity and quality of implementation of the intervention adversely or supported it.

One approach to quantifying implementation characteristics is to create an index such as a fidelity index comprised of a checklist of core program elements for which each site would receive a score denoting the degree of adherence to the program model. Composites of variables in each of the key areas of interest (e.g., quality of service, adaptation) could be created to summarize level of implementation at the individual site level and aggregated to the replication site level. These same variables would then be used in analyses linking implementation to program impact (as described in the following section).

The final step in the explanatory analyses is to attempt to link variations in aspects of implementation such as fidelity to the program model and quality of the services provided, as well as other factors, to participant impacts. As with the analyses that precede them, the primary approach is a qualitative one, although we will explore the potential of methods such as performance analysis (Mead, 2003, cited in Werner, 2004) to model the relationships between program activities and processes and outcomes.

Quantitative Approach to Linking Implementation to Program Impact

Researchers and policymakers are increasingly interested in using quantitative methods to link variation in program implementation to variations in the impact of the programs on participants (Zvoch, 2012). While the analytic strategies are relatively new and still being developed, it is our hope that it will be possible to supplement and strengthen the qualitative analyses with a more quantitative approach.

Of course, such a strategy will be feasible only if the impact study determines that there are variations in program impact as well as variations in program implementation. While it is likely that such variation will appear at the site level, these analyses require that we observe variation at a lower level – at the health educator, clinic or school level, so that there are a sufficient number of units to make the analysis feasible.[6]

We propose to attempt to link key dimensions of implementation to program impact at two levels of policy interest. First, within each program model, we will examine relationships between key aspects of implementation and program impact. This analytic strategy enables us to leverage power by pooling data from relatively small studies in each multi-site replication in order to enhance our ability to detect relationships or even possibly explain variation in program impact of an individual model. Secondly, we will explore the possibility of using pooled data from all three multi-experiment replications to yield similar kinds of information about the TPP replications overall.

Our approach to these analyses borrows from the work of Bloom, Hill, and Riccio (2003) and others (e.g. Greenberg et al., 2003) who have used the same approach in linking implementation and effectiveness. Although their topic was welfare-to-work and their sample was considerably larger, their basic approach is promising, and the current study provides sufficient numbers of rigorously designed experiments to support exploratory analysis of links between key aspects of implementation and program impacts.

As noted above, our model posits that fidelity of implementation, quality of service delivery, adaptation and participant responsiveness are important aspects of implementation over which programs have some control, in which we could expect some variation, and which prior research has linked to program outcomes. These four key components of implementation are candidates for use in the analysis, although in all likelihood, three or fewer will actually be used.

We would expect the greatest amount of variation to be at the sub-site or performance site level (such as a clinic, in a replication with multiple clinics, or a school or other setting). Our approach then entails three steps. First, we will construct a small number of quantitative measures of implementation. Second, we will use measured program impacts (as described above, in the impact study). Third, we will use multilevel modeling (Raudenbush & Bryk, 2001) in which participants (level 1) are grouped by performance site/sub-site (level 2), which are in turn grouped within sites (grantee site) to explore the relationship between program implementation and short-term and longer-term TPP program impacts.

---

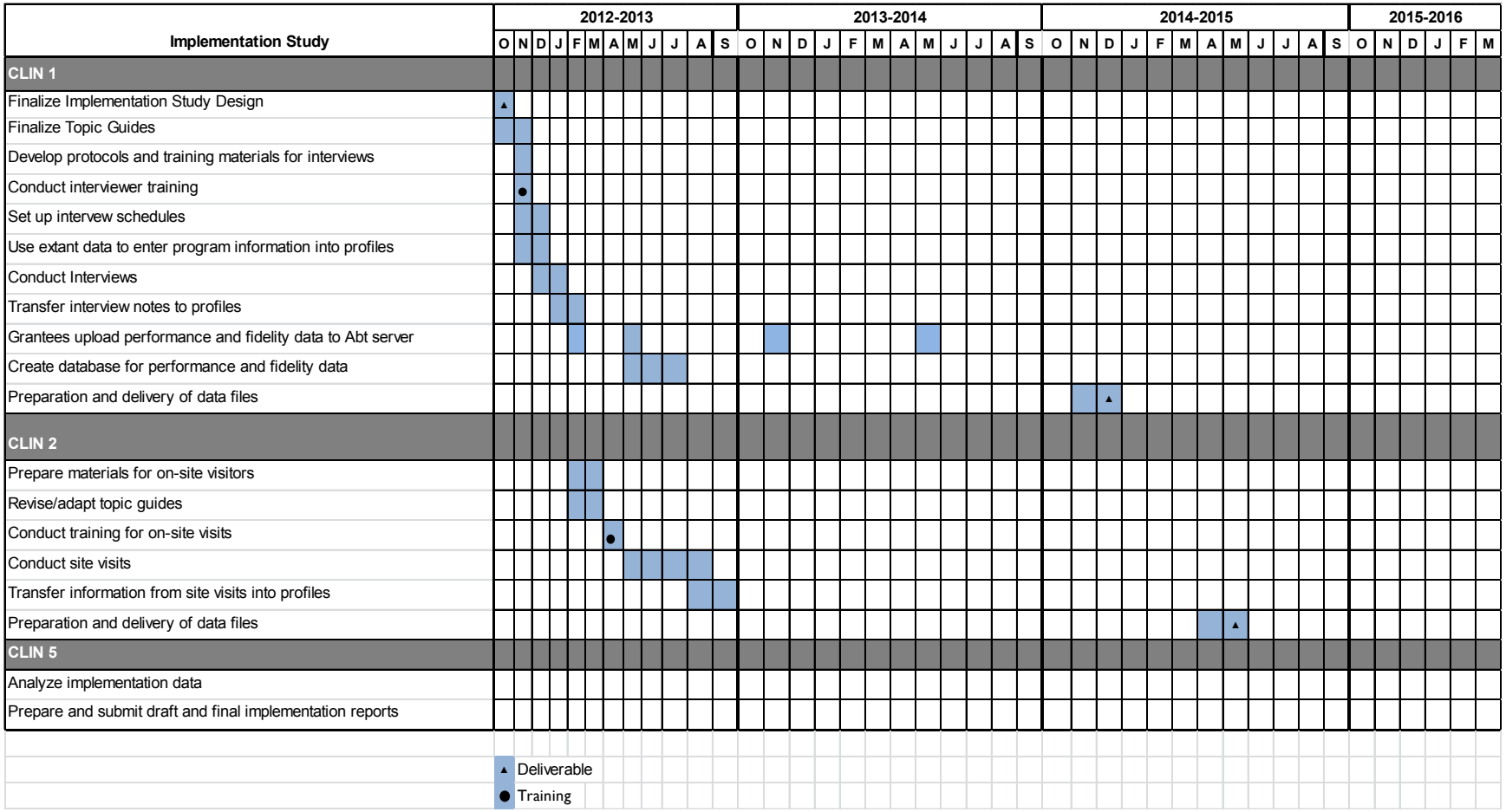
[6] Before proceeding with an analysis of the impact of program characteristics, we will determine whether the differences in program effects are statistically significant. If they're not, we will not estimate how program impacts vary with program characteristics. To make this determination, we will use a chi-squared test to test the null hypothesis that the effects are identical across sites.

The key dimensions of implementation that we would propose using are: administrative and supervisory supports; fidelity to the program model; quality of service; adaptation; and participant responsiveness. We would use the fidelity index for each model described above (descriptive analysis). The quality of service and responsiveness measures will draw on observations and focus group data, and the adaptation data will be drawn largely from program documents and interviews with program and frontline staff. This approach enables us to explain the variation in experimental impact findings by isolating the independent influences on it of the implementation factors of interest.

Timeline for Study Activities

Exhibit 4 displays the timeline for implementations study activities, by service component. Schedules for the analytic and reporting activities (in CLIN5) remain to be negotiated.

Exhibit 4: Implementation Study Timeline for Activities



References

Berkel, C., Mauricio, A., Schoenfelder, E., & Sandler, I. 2011. Putting the pieces together: An integrated model of program implementation. *Prevention Science* 12, 23-33.

Bloom, H., Hill, C., Riccio, J. 2003. Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management* 22(4), 551-575.

Dane, A. & Schneider, B. 1998. Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review* 18, 23-45.

Durlak, J. & DuPre, E. 2008. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology* 41, 327-350.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. 2003. A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research* 18, 237-256.

Fixsen, D., Panzano, P., Naoom, S. & Blase, K. 2008. Measures of implementation components of the national implementation research network frameworks. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, the National Implementation Research Network.

Granger, R.G. 2011. The big why: a learning agenda for the scale-up movement. *Pathways*, Winter 2011, 28-32.

Greenberg, D., Meyer, R., Michalopoulos, C., & Wiseman, M. 2003. Explaining variation in the effects of welfare-to-work programs. *Evaluation Review* 27(4), 359-384.

Mead, L. 2003. Performance analysis. In MC Lennon, & T. Corbett (Eds.). *Policy into Action: Implementation Research and Welfare Reform*. (107-144). Washington, DC: Urban Institute Press.

Raudenbush, S. & Bryk, A. 2002. *Hierarchical linear models: Applications and data analysis methods, 2nd ed.* Thousand Oaks, CA: Sage Publications.

Zvoch, K. (2012). How does fideligy of implementation matter: Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. *American Journal of Evaluation* 33(4), 547-565.