CHAPTER 16

Building Evaluation Into Your Work

Why evaluation is important in the field of teen pregnancy prevention

How could we have worked so hard on teen pregnancy prevention and yet know so little about how to solve this problem? We have lots of good ideas, "best bets," and promising programs but little unassailable data on what effects they have. One of the reasons for this is our failure to document the results or even the content of most programs. Satisfied with good intentions or a few stories about program successes, many efforts are continuously funded without any strong evidence that shows they work.

Several authors have recently summarized what we have learned about how to prevent teen pregnancy (see box). Each of them has found few documented successful programs and has lamented the sparse and fragile state of available research. Neglect of evaluation in teen pregnancy prevention has left dedicated program directors with little on which to

The state of program evaluation in the teen pregnancy field

"Although many attempts have been made to alter adolescent sexual and childbearing behavior...most of these undertakings have not been accompanied by rigorously conducted evaluations. Thus, anecdotes describing 'successful' programs are often followed by unpublished reports that may or may not confirm that a 'significant' change occurred in the behavior of program participants....Moreover, the paucity of information...[has] led to an unwarranted sense of pessimism..." (Frost & Forrest, 1995, p. 188).

"The fact that only 23 evaluated programs surfaced for detailed review merits comment....many obstacles stand in the way of conducting strong program evaluation: (1) cost, (2) methodological difficulties, and (3) a social environment in which research on fertility-related topics may be seen as controversial" (Brown & Eisenberg, 1995, p. 228-229).

"Regrettably, aside from a few well-designed, well-implemented, and well-evaluated studies, most of the programs that have been implemented to affect teenage childbearing have been small, ad hoc, and poorly designed short-term projects lacking a useful evaluation strategy" (Moore et al., 1995, p. viii).

pattern their own efforts. This has consequences, in turn, because repeated failures to curb teen pregnancy can give an impatient nation the impression that nothing works.

Why you should evaluate your prevention program

There are three main reasons why you should consistently incorporate evaluation into your program planning and budgeting:

Good evaluation leads to good program management. A program director can better direct a program when it is clear exactly what is being delivered, whether it is reaching the intended targets, and what results are occurring. A regular flow of data on these vital topics can enhance the impact and smooth functioning of a program.

RESEARCH NOTES

The hallmarks of good evaluation

Measure behavior, not just attitudes and beliefs—especially because attitudes and intentions about sexual issues are often unrelated to actual sexual and contraceptive behavior.

Conduct long-term follow-up (of at least one year). Short-term results do not necessarily translate into long-term results, and some effects are not apparent in the short-term.

Conduct proper statistical analyses—which means, among other things, stating the hypothesis you are testing before collecting the data.

Include a sufficiently large sample size. While adequate sample sizes may vary, experience suggests that a sample of at least 1,000 subjects may be advisable, especially if the intervention has a modest effect. However, programs with fewer subjects can still learn from evaluation.

Share both positive and negative results. When negative evaluations are not shared, the meaning of positive results are compromised.

Replicate studies of successful programs. A single positive evaluation may be the result of a particular instructor, a particularly successful group of youths, or other factors in the community. Replication and evaluation in other communities and settings must be done.

Adapted from Kirby, 1997

3 GOOD REASONS TO EVALUATE YOUR PROGRAM

- **it leads to good program management**
- **it prevents resources from being wasted on things that don't work**
- **it helps builds a strong scientific base for the field**

Good evaluation prevents resources from being wasted on things that do not work. The history of efforts to reduce teen pregnancy prevention is littered with interventions that were too little, too late, too brief, and too weak. While we decry the lack of resources for teen pregnancy prevention, many dollars can be wasted on programs that cannot succeed and that, unknown and undocumented, fail.

Good evaluation helps to build the scientific knowledge base for the field of teen pregnancy prevention. Before a new drug becomes available to the public, it is carefully tested using the highest scientific standards. Indeed, the American public expects no less. In teen pregnancy prevention and in other human service programs, by contrast, we have let untested programs occupy valuable school hours and critical after-school time. Without any data to show that what we are doing works, or even that it does no harm, we dispense untested interventions. Surely we can find data-based, well-tested solutions to this problem if we evaluate what we are doing, discard what is ineffective, and put our collective will and our resources behind what does work.

Getting help with evaluation

If at all possible, involve a trained evaluator in your program when you are in the planning stages. Even if the program has existed for some time, an evaluator can help.

Choosing an evaluator

There are four primary sources for professional evaluators:

- colleges and universities;
- consultants who are skilled in evaluation;
- evaluation firms; and
- in-house evaluators who become employees of your program or agency.

Regardless of which source you use for your evaluator, look for

Involve an evaluator early—in the program planning stages.

someone who is trained in research methodology and statistics. At colleges or universities, try the sociology or psychology department, the school of public health, or a research institute that studies social issues. University faculty may have consultant days they can use to do your work or they may ask you to buy some portion of their time. There are also individual consultants and firms who do evaluation work. The difference in these two is whether there are other staff (a firm) or just an individual consultant doing this work.

Larger programs sometimes hire full-time employees to take care of their evaluation work. While very senior people are usually not available for this role, master's level researchers often are. Some people would argue that outside, independent evaluations have greater credibility than those done by employees, but others would say that there is no reason to suspect the integrity of an in-house evaluator. If evaluators can be swayed by program staff to shade the outcomes in a report, why not suspect that they could also be swayed by funders who want their programs to have positive results? Most evaluation reports are read not only by program audiences but by other evaluators as well, and published evaluations usually undergo careful review by well-trained experts. These safeguards keep most evaluators objective, regardless of who pays them.

One of the most important things to consider in hiring an evaluator is experience with program evaluation. This work differs from academic research, not in its methods, but often in its pace, its real-world context, and in how "pure" it can be. Make sure your evaluator has done research in the program world in which you work.

Ask to see a potential evaluator's resume as well as a recent evaluation report in which he or she has had a major role. Can you understand it? If not, perhaps this is not your person. Make sure this evaluator can

KEYS TO SUCCESS IN HIRING AN EVALUATOR

- **look for experience with program evaluation—preferably with teen pregnancy prevention programs**
- **make sure that the evaluator can write readable reports**
- **check on the evaluator's time availability**
- **ask about costs, including overhead costs**
- **ask who will do which tasks of the evaluation**
- **agree on ownership of the data and publication rights and conditions**

write a report that will be useful for audiences who lack technical training.

Find out about the time your potential evaluator has available for your project and who will actually do your work. Is a professor going to hand off the work to a student? Is a consultant going to give the work to junior staff? Both of these strategies may be okay, but it is best if you are not surprised. If you are hiring an individual rather than a firm, it is especially important to know who will do the clerical tasks associated with the evaluation. Volunteers, students, or clerical personnel can often collect data, make telephone calls, or enter data into a computer. You do not want to pay Ph.D. rates for this kind of work.

Make a clear agreement with your evaluator about the confidentiality and ownership of data. The media or your funder may call to ask your evaluator about your results. Be sure that your evaluator understands that your data are not to be discussed with anyone without your permission. If you are hiring an evaluator who is interested in publishing your results, agree in advance on how this will be handled. Who will be the co-authors? Do you want to review and approve reports or articles before they are submitted? Remember you are not hiring a public relations firm. You are hiring an evaluator and there is always the chance that the results will not be what you had hoped.

Whether your evaluator must be located in your community depends on several things. If the evaluation design will require frequent on-site presence of the evaluator, travel costs may become prohibitive if your evaluator is far away. On the other hand, you should also consider "best capacity" as well as "local capacity." Some evaluators in the teen pregnancy area have vast national experience. They have existing instruments and procedures that may be useful to you and they have seen many teen pregnancy programs. The travel costs may be worth what you get back in expertise.

Evaluation doesn't have to be scary and threatening. It becomes less so when you choose an evaluator whom you like and trust. It is good to look for an evaluator who has good interpersonal as well as research skills. Remember, never accept an evaluation design that you do not understand or that does not feel right to you, no matter who suggests it.

THE BOTTOM LINE...

Never accept an evaluation design that you do not understand or that does not feel right to you, no matter who suggests it.

REMEMBER...

the program that adds relatively small amounts of evaluation money to each proposal soon has a good-sized nest egg.

Budgeting for evaluation

The demand for evaluation is perhaps a little ahead of the understanding of its costs, but many requests for program proposals now require a detailed evaluation budget, and most funders asking for this work also pay for it. Programs often add dollars to their program proposals to pay for their rent, their accountants, their personnel managers, and other standard expenses. It is easy to make the case that evaluation is one such cost. The program that adds relatively small amounts of evaluation money to each proposal soon has a good-sized nest egg.

Each evaluation carries its own costs, so there is no simple formula for deciding how much to spend. It all depends on what you plan to do. If you plan to track program clients over the long-term, particularly if these clients are highly mobile, your costs will go up. If your program needs monthly rather than annual reports, the costs will go up. Senior level evaluators are more expensive than are less experienced evaluators. The table on the opposite page shows items that generally appear in evaluation budgets and some variations in their cost ranges. Of course, your evaluation spending will have to fit within your overall budget.

Designing your evaluation plan

Many requests for program proposals now require a detailed evaluation plan. Evaluators can help you create this plan and some will do this without charge, if you make a commitment to hire them as your evaluator if the project is funded. Find an evaluator who will work with you this way so that you can consider the appropriate design and costs for this work right at the beginning of your project. You should find that the concrete thinking habits of evaluators and their experience with other programs like yours will be helpful to your program design.

Your plan should include two types of evaluation:

- process evaluation; and
- outcome evaluation.

Most programs routinely collect information that can be used for process evaluation. Outcome evaluation is more demanding

Components of an evaluation budget

Evaluation Line Item	Range of Costs
The evaluator	
This person plans your overall design and takes responsibility for the product.	Costs range from about $300 to $1,000 per day.
Data collection costs or incentives	
Sometimes there are costs for paying interviewers or for buying incentives, such as movie passes to give to young people.	Interviewers are usually paid by the hour (usually $10-12 per hour) or by the completed interview; calculate incentive costs per head.
Data entry	
Input of data to computer, if you are not doing this yourself. Number of forms per hour x cost per hour.	Costs range from about $5 to $15 per hour.
Data analysis	
The person who does the actual "number crunching"; could be the senior evaluator or could be another person (be sure to ask because it affects costs).	Costs per hour or day ranges are very broad, from graduate student to full-time employee wages, but estimate $10 to $30 per hour.
Supplies	
Computer supplies, paper, report covers, and the like.	Varies.
Copying	Costs range from 5 cents to 20 cents per page.
Telephone/fax/e-mail	Nonlocal evaluators will increase these costs. Include any costs for telephone interviews or surveys.
Postage	
Normal postage costs, plus costs for mail surveys, if you conduct them.	Varies. If you are gathering any data by mail, do not forget to add enough for return postage.
Travel	
Both local and nonlocal evaluators have some travel associated with work on your project.	Nonlocal evaluators cost more in this category, especially if you need a lot of meetings.
TOTAL DIRECT COSTS	SUM OF ABOVE
Administrative Costs (or overhead)	This pays for a portion of the evaluator's rent, heat, etc. It is sometimes figured on total budget, sometimes on salary and wages only, and sometimes it is hidden in the budget. It is a legitimate expense, regardless of how it is presented. If the evaluator is already on staff, remember that you still need to account for these costs.
TOTAL COSTS	DIRECT PLUS ADMINISTRATIVE

and requires a different type of expertise, so fewer programs conduct this type of evaluation. However, both are important if you are to get a true picture of whether your program is achieving what it set out to do. Both of these types of evaluation are described more fully later in this chapter.

Preparing your program for evaluation

Once you have developed a plan for evaluating your program that takes into account your available staff and budgetary resources, you will need to complete at least two important tasks before you actually begin evaluation activities:

- decide on your target population; and
- create a program model.

Decide on your target population

Early sexual activity, non-use of contraception, and early pregnancy are more likely to occur among some youth than among others. One of the most important predictors of these behaviors, for example, is poverty (Kirby, 1997). The most efficient programs serve those young people most at risk of pregnancy and do not expend resources on young people who are the least likely to get pregnant. Indeed, one of the most basic evaluation questions to be asked is whether the program reaches those who are most at risk.

Create a program model

A truck driver who sets off for a destination has a road map. She does not drive aimlessly around for several days, expecting to suddenly arrive at her destination. Instead, using her map as a guide, she follows road signs and expects that two right turns, 500 miles on the Interstate, and exit 32 will lead her right where she wants to go. Program personnel would do well to emulate this model, preparing the equivalent of a roadmap to their destinations.

In evaluation terms, this roadmap is a "program model," also called a logic model or a theory of change. A program model specifies:

- *your planned strategies, activities, or interventions —*

BEFORE BEGINNING ANY EVALUATION, COMPLETE TWO IMPORTANT TASKS

- **decide on your target population**
- **create a program model**

sometimes called process goals by evaluators, this portion of the model should state exactly what will be delivered to whom and by when;

- *the short-term results expected* from these activities or interventions; these outcome or results goals should state exactly what will happen in the relative short term to young people, their parents, the community, service providers, or others you plan to reach; and
- *the longer-term outcomes expected* as a result of both the program's activities and its short-term outcomes.

A program model moves program staff from "vague-speak" to a clear statement of what they are going to do and what their actions will produce. Here are some examples of "vague-speak" that do *not* lend themselves to good evaluation:

- We are going to develop youth to their fullest potential.
- We are going to provide quality care for all our patients.

Why are these goals so frustrating to evaluators? Because they lack concrete definition and thus allow wide interpretations. How would we know one of these fully developed youths if we saw one? What exactly do we mean by quality care?

While some can describe what they really mean by this kind of talk, it may also mask confusion or disagreement about program goals. Do youth developed to their fullest potential do well in school? Do they use drugs? Do they use contraception if they have sexual intercourse?

Creating a program model has several immediate benefits:

It makes the interventions and outcomes very clear. A changing list of funders and the normally high staff turnover in human services make it easy for a program to lose its way. Every program should have a clear program model, posted where everyone can see it, so that new employees, the press of day-to-day work, and tempting but tangential offers from potential funders do not blur its focus.

It protects your program from inappropriate or excessive expectations. A program with such a model is not likely to be

THE 3 COMPONENTS OF A PROGRAM MODEL

- **strategies, activities, interventions**
- **anticipated short-term results**
- **anticipated long-term results**

PROGRAM MODELS ARE ALSO CALLED

- **logic models**
- **theories of change**

BENEFITS OF A PROGRAM MODEL

- **it makes the interventions and outcomes very clear**
- **it protects your program from inappropriate or excessive expectations**
- **it enables you to define measurable outcomes**
- **it checks the logic of your assumptions**

held accountable for inappropriate outcomes by its funders, its board, outside evaluators, or any other group. When a program takes the initiative to define its own intended results, it is less vulnerable to misunderstanding.

It enables you to define measurable results. With clarity comes measurability. Programs with such models can more easily talk to evaluators about exactly what measures will be appropriate.

It checks the logic of your assumptions. So often in teen pregnancy prevention programs, we have designed interventions that are not equal to the task. Programs plan to undo 15 years of poverty, lack of aspirations, and poor role modeling with a 45-minute curriculum. Programs provide contraceptive services or abstinence education in only one of five high schools but expect a drop in the county

Sample program model

Program Processes, Activities, Strategies, or Interventions	Expected Short-term Outcomes or Results	Expected Longer-term Outcomes or Results
By June of this school year, 150 teens aged 13 to 15 in the local community will have: • received after-school tutoring and homework help; • participated in a career education program; • taken part in a group entrepreneurial program to produce and sell a product; • been explicitly encouraged to delay sexual activity; • received contraceptive services if sexually active; • taken part in community service; and • taken part in a recreational program, including field trips to college campuses and other destinations designed to expand their horizons.	*These young people will show an increase in:* • homework completion; • school attendance; • later ages of first intercourse; • contraceptive use; and • career aspirations.	*By the end of the program's second year, these young people will show:* • lower rates of pregnancy. NOTE: This program may well have additional longer-term results related to education, careers, and youth development, but the outcomes of greatest interest to you will be those that help you achieve your overall goal—reducing teen pregnancy.

pregnancy rate. A clear program model enables us to check the underlying reasoning and perhaps avoid these costly errors.

The chart on the opposite page shows a sample program model for a teen pregnancy prevention program, including program activities and expected results. The long-term outcomes are usually the project's major objectives and depend on the success of the program's activities and any short-term changes made by clients that result from them.

Designing your process evaluation

Process evaluation measures what your program provides and the characteristics of those who receive it. It has several important purposes:

To ensure that services are delivered as planned. There are many ways for services to be derailed. Staff become too busy to spend as much time mentoring as was planned. Some portion of the curriculum is never delivered because teachers feel untrained on a topic. Home visits do not occur because the neighborhood appears too dangerous. Staff avoid the subject of sexual behavior because they are uncomfortable discussing this subject with teens.

A good process evaluation would discover these problems early and function like a smoke detector. The "Field Notes" on the following page tell the story of a program that was not really happening, and thus, was doomed to fail. A simple process evaluation discovered this problem and allowed the program to correct itself.

To document whether the program is reaching its target population. Young people are not at equal risk of having intercourse, of not using contraception, or of getting pregnant or causing a pregnancy. Programs sometimes deliver their interventions to any available group and may thus reach only those at a relatively low risk of these behaviors or some group other than those initially targeted for attention.

To detect and monitor variations in effectiveness within the program. Programs often have higher success rates with some young people than others. Perhaps younger teens have higher attendance rates than older ones. Perhaps the boys hear

and act on certain messages but the girls do not. Data on attendance and characteristics of young people can be combined with outcome data to explore with whom the program is most and least successful. Such information can be used to adjust the program or to limit its outreach to those with whom it is most successful.

There are many tools to collect process evaluation data, including clinic intake forms and attendance sheets. These tools are discussed in detail later in this chapter.

FIELD NOTES

Process evaluation saves the day

A school-based clinic gathered information on all its new patients, including whether they had ever had intercourse and the age at first intercourse. After these data were collected for some time, the following findings emerged:

- 183 students had reported that they had had sexual intercourse;
- 108 of these had their date of first intercourse recorded in the database.

Why were some of the data missing? Did the young people not know the answer to the question? Did they not understand the question? Did staff always ask the question? A further investigation showed that:

- 704 young people had been into the clinic three or more times and had not yet been asked whether they had ever had sexual intercourse.

How could this be? An initial meeting of the clinic staff to discuss this resulted in an argument. The pediatric nurse practitioner said it was the social worker who did not ask questions properly, while the social worker believed it was the medical staff causing this problem.

After some time, one staff member said, "But they are so young. They will be embarrassed if I ask them about sex." If this teen pregnancy prevention program had staff members who were reluctant to talk about sex, what chance did it have for success?

Group work and staff training were chosen as the solutions for this problem, but the problem never would have been discovered if it had not been for the process evaluation.

Designing your outcome evaluation

An outcome evaluation answers two questions:

- Were there any changes among those served?
- Did the program cause those changes?

These are not easy questions to answer, and some oft-used strategies do not provide these answers. For example, client testimonials about how much they like the program, staff reports of how many hours they work, poignant vignettes about high-profile clients, and even counts of successes and failures are not enough to answer both of these questions. The chart below shows examples of what not to do in evaluating outcomes.

Strategies that, by themselves, do not answer outcome questions

Client testimonials about how much they like program staff or their services
Someone asks a program director if her program works. "Why, yes," she replies. "Over 90 percent of our clients say they are very satisfied with our services."

We dislike so many things that are good for us and we love so many things that are not. If they also like you or your program, does that really tell us anything about what *impact* you are having?

Staff reports about how hard they work
Someone else asks a program director if her program works. "Why yes," she replies. "How do you know?" her questioner persists. "Why we come over here and start work at 6 a.m. Often we do not leave until 10 at night!" Does that say anything about program results? Unfortunately, it does not. Perhaps the sad truth is that this staff is inefficient or spending too much time on the wrong things.

The poignant vignette
Yet another curious person says, "Does your program work?" "Oh yes," replies the eager program director. "How do you know?" her questioner persists.

"Well, let me tell you about Jennifer. She was a school dropout and having sex with numerous partners without any protection. We worked with her for six months and now she is enrolled in college and has given up having sex for the moment!"

These are good stories and you should collect them. But the real question for a good outcome evaluation is how many such stories there are. No one will ever say that a program is not working and illustrate it with a story about a client who failed.

It is important to add that not all programs are ready or able to do outcome evaluation. Some, for example, are too new and small; others are unstable, constantly changing their approach, making it almost impossible to define the precise intervention or figure out its effects. And some programs cannot afford to spend what it takes to do top-flight outcome evaluation. If fact, some have suggested that *funders* only require *some* of the programs they support to do outcome evaluation—those that are sufficiently large, stable, and well-managed. Nevertheless, measuring the results of a program remains enormously important, even if done on a very modest level.

There are a number of ways to approach outcome evaluation. Several are given below, starting with the most rigorous.

True experiments with random assignment

Random assignment is a research design with a long history in science. In essence, it means setting up two groups by dividing all possible clients on a random basis; one group gets the program intervention and the other does not. It is the technique that will answer the question of whether there were any changes in the group served, and whether the program caused these changes. It is the strongest design available for getting at cause and effect. The table below shows what this classic experimental design looks like.

The classic experimental design

Randomly assign individuals into two groups:

The Program Group	Take a "before" or baseline measure of outcomes	Give the group the program, treatment, medicine, or whatever is being tested (the "intervention")	Take an "after" or follow-up measure of outcomes
The Control Group	Take a "before" or baseline measure of outcomes	Do not give these people the intervention	Take an "after" or follow-up measure of outcomes

Compute the effect of the program as equal to:

Changes made by program group minus changes made in the control group.

Subtracting changes made by the control group takes away changes that would have occurred anyway because clients got older, wiser, or healthier, changed their habits, or because of anything else that might have happened without the intervention.

STEPS IN THE CLASSIC EXPERIMENTAL DESIGN

- **randomly assign individuals into one of the two groups**
- **take baseline measures of both groups before intervention starts**
- **monitor both groups**
- **take follow-up measures of both groups after intervention ends**
- **compare outcomes for both groups**

In a teen pregnancy program, you would plan to carry out such a design as follows:

Randomly assign the pool of those recruited, assigned to, or wanting to be part of your program to one of the two groups. You could put their names in a hat and draw out half of them to be assigned to the program. Those whose names remain in the hat become "the control group." Another way to assign clients randomly is to alphabetize all the names of the applicants and choose every other name on the list for the program and control groups.

The point of this procedure is to ensure that the young people in the two groups are as much alike as possible, differing only in their receipt of the program. If you let the most motivated or the most needy youth join the program and use those left over for the control group, the two groups are different from the outset. If you later find that the program is getting good results, it may be because of the kind of kids who joined in the first place, not because of the program itself.

Some may wonder whether this procedure is unethical. Certainly it is when the intervention is something that we already know works. For example, we would not want to give a random half of clinic patients seeking contraception a birth control pill and the other half a sugar pill. The experiments on birth control pills have already been done and we know they work when used properly and consistently. Immunization is another well-tested intervention.

How does this general idea apply to your program? Do you know that what you are doing works because it has been rigorously tested? If not, you may be enrolling young people into a program that is not helping them and, perhaps, keeping them from participating in other activities that are more effective. In fact, you could be doing harm. The point is that you should probably not rule out this strong research design without carefully considering its use.

Some programs use random assignment to test variations on their programs. For example, half of the young people may get one kind of counseling, while a random other half get another. Perhaps half of the young people

may receive "business as usual" while the other half receive a new program. In these designs, none of the young people is denied services.

There is no need for the program and control groups to be the same size. If you have a large pool of young people from which to assemble your evaluation groups, more of them could go into the program group, provided that you have enough in each group to detect significant outcomes, should they occur. An evaluator or statistician can help you figure out the minimum numbers necessarily. The calculation has to take into account how much change your program can be expected to produce.

Take baseline measures in your program and control groups during the first few weeks of the program. To measure whether your program has any impact on particular behaviors, it is important to first assess the behaviors *before* you begin work with your program clients. Therefore, you are likely to want so-called "baseline measures" of, for example, whether teens are having intercourse, at what age this began, whether, how often, and what kind of contraception they are using, whether they have ever been pregnant or caused a pregnancy, and what their knowledge, attitudes, and skills are (see the next section of this chapter for more on measuring these behaviors and attributes). Of course, if your program is focused on different outcomes, like parent-child communication rather than sexual behavior, the factors measured at baseline will be very different.

Besides getting these measures, ask the young people in the control group whether they are part of any other programs in your community. Unless you are the only teen pregnancy prevention in town, some of your control group members may be getting services or interventions like those you offer. You may find that you end up testing how strong your program is compared to other programs rather than the effectiveness of your program compared to no services at all.

Keep track of participation in your program. The reasons for doing this for process evaluation purposes has been discussed already. In an experimental design, this information is valuable because you might want to later look at results for young people who participated more regularly and for those who participated less.

Take follow-up measures from your program and control groups. The follow-up measures of outcome are usually identical to the baseline measures. After young people have been in your program for a year or longer, how many

are having intercourse, how many are using contraception, and how many have been pregnant or caused a pregnancy? If you are testing a short-term intervention, such as a program to change teens' knowledge, the baseline and follow-up measures may occur close together, being only weeks or even days apart. Still, we would expect the changes in knowledge to be greater and, of course, more positive among the program group than among the control group.

Look at your post-program results for both the program and control groups. The appropriate analysis for most of these outcomes is to take into account where each group began or their baseline measures, and then compare where they were at the end of the program. This is not the same as comparing only the post-test or follow-up levels of these outcomes between the two groups.

There are some outcomes, such as one-time events, for which "before" measures are not possible. Examples of such outcomes are age at first intercourse and dropping out of school. For these outcomes, it is appropriate to compare just the follow-up measures or program exit measures between your program and control group youth.

WHAT IS THE APPROPRIATE ANALYSIS FOR OUTCOME MEASURES?

Compare the before-to-after changes between the two groups, not just the difference in "after" levels between the two groups.

Whenever changes occur among program clients, you will need some way to decide whether those changes are important. What if a group of young people raised their knowledge scores by 10 percent? Is that change significant or important? To decide the answer to this question, researchers use tests of "significance." These tests determine when a particular change is likely to be due to something *other than* chance. For example, if we flipped a coin 100 times and we got 90 heads, we would suspect that we did not have a fair coin and that something else is operating here besides chance. Tests of significance help us make these decisions for findings about change or about differences.

You should know that statistical significance is strongly affected by sample size. In very small samples, changes will have to be very large to be significant. In large samples, very small changes will be statistically significant, even if these changes are not practically important. This is one of the reasons that small programs sometimes decline to do outcome evaluation; they simply do not have enough participants to generate

WHEN IS A BEHAVIOR CHANGE SIGNIFICANT? THE ANSWER DEPENDS ON SAMPLE SIZE.

- **in very small samples, changes have to be very large to be significant**
- **in large samples, very small changes will be statistically significant, even if these changes are not practically important**

changes large enough, or common enough, to show up in statistical analyses.

The following table shows a simplified way to calculate relative change in outcomes. Your evaluator or a statistician can provide help in using more sophisticated techniques, such as regression analysis, for this purpose.

Other designs

The classic experimental design outlined in the previous section is the most powerful strategy available for answering the two outcome evaluation questions: Were there any changes among those served? Did the program cause those changes?

Some programs, however, lack the resources to carry out such a demanding design. Sometimes it is premature to use a random assignment experiment if a program has not yet done any evaluation at all. In these circumstances, a program may choose an alternative approach to evaluation. Although all have less ability to assure that a given program was the *cause* of any change, programs will be able to learn at least something about their program's impact.

Using a comparison group

If random assignment is not feasible, perhaps an evaluation can at least find another similar group of young people to serve as comparisons. Of course, unless these young people tried to get into the program being evaluated, they are likely to be different in some way from those who did apply. If they are about the same age, ethnicity, gender, and socio-economic status as those in the program,

Hypothetical data from a classical experimental design

Group	PRE-TEST knowledge scores was:	POST-TEST knowledge scores was:
Program group	20 percent correct	60 percent correct
Control group	20 percent correct	40 percent correct

To determine the NET effect of the program...

compute the changes made by the program group MINUS the changes made by the control group, or:

The Program Effect = (60 - 20) - (40 - 20) = 40 - 20 = 20, or a 20 percentage point knowledge gain. Tests of significance could then be used.

though, they might make a good comparison group.

Sometimes, other schools, other communities, other neighborhoods, or even other programs can provide a comparison group of young people for your evaluation. The important thing is for your comparison group to match your program group in as many ways as possible. Those who drop out of your program do not make good comparisons since they are clearly different from those who stay in.

If you use a comparison group, the steps are the same:

- take baseline measures;
- keep track of what services are received; and
- take follow-up measures and look at changes in your program group versus changes in your comparison group.

Using available data

Some programs will find it difficult to follow either control or comparison groups because they lack the staff or funds to pay for this activity. If this applies to your program, you might at least secure some printed or available data on behaviors with which to compare your program group. For example, what is the teen birth rate in your county? Do your program teens have a lower rate?

Although teen birth data are relatively easy to come by, teen pregnancy data can be harder to get because a true pregnancy rate includes births, abortions, and miscarriages and counts of the latter two are sometimes inaccurate (see below). Comparison data for knowledge, attitude, or skill changes are less available, but a search of research literature might produce data from another evaluated program. A good project for a graduate student would be to assemble relevant comparison data (preferably local data) that you might use for comparison purposes in your program's evaluation. This compendium could then be updated every three years or so.

Be careful to choose a comparison number that comes from a population as close to yours as possible. If you are serving very high-risk young people, for example, the overall teen pregnancy rate for the United States may not be a good comparison because that rate includes

WHAT'S THE DIFFERENCE BETWEEN A CONTROL GROUP AND A COMPARISON GROUP?

- **A *control group* is one produced by random assignment.**
- **A *comparison group* is similar to the intervention group, but not created by random assignment.**

advantaged children who are at less risk than those you may serve. Of course, if your program clients have lower pregnancy rates than do national teenagers, your program looks good. If their rates are higher, this comparison is less useful.

This strategy is not, of course, as credible as the other designs described here. It is easy to choose an inappropriate comparison standard or to not know how appropriate an available number might be. Still, having some comparison data can show that a program looks promising and thus make it easier to attract funding for a more credible evaluation. Again, a trained evaluator can help a lot.

Measuring outcomes and variables

Teen pregnancy prevention programs are interested in one major outcome—lower rates of pregnancy. Variations or changes in two behaviors can lead to this outcome:

- sexual intercourse; and
- contraceptive use (or lack thereof).

These behaviors can be considered direct influences (often called "proximal variables") upon the principal outcome of pregnancy. In turn, proximal variables may be influenced by other factors (often called "antecedent variables"), including socioeconomic status, knowledge, attitudes, skills, and personality characteristics. Most teen pregnancy prevention programs focus on changing antecedent variables (particularly knowledge, skills, and attitudes)—with the ultimate goal of reducing sexual activity and/or increasing contraceptive use among teens, thereby reducing teen pregnancy rates.

The baseline and follow-up measures that you take during the course of your program will allow you to track changes in antecedent and proximal variables. In addition, some programs also measure related indicators, including school behavior, drug use, and involvement in the juvenile justice system.

Measuring the major outcome variable—teen pregnancy

Teen pregnancy is generally measured either by using available data for an area or by asking program clients about their

pregnancy histories. Available data on pregnancies usually lag some years behind their actual occurrence because they require combining counts of births, abortions, miscarriages, and stillbirths. Counts of these events are kept by different agencies and institutions. Data on teen births are much easier to get than data on teen pregnancies. In general, data for smaller geographic areas, such as zip codes, are harder to obtain than data for larger areas, such as states. Still, this information is available.

Asking young clients about their pregnancies is straightforward, though typically requires tact, sensitivity, and confidentiality. Girls can be asked if they have ever been pregnant and boys can be asked if they have ever caused a pregnancy. It is also possible to ask when each pregnancy ended and how it ended (live birth, miscarriage, abortion, or stillbirth). If you are gathering data to evaluate the impact of your program on pregnancies, be careful to get the date of each pregnancy so you can tell if it occurred before or after enrollment in your program. At baseline, you might ask young people if they have ever been pregnant and then, at each subsequent follow-up assessment, ask about pregnancies for that interval (e.g., within the last six months, during this school year).

Measuring proximal variables

Each of pregnancy's proximal variables—sexual intercourse and contraceptive use (or lack thereof)—has several dimensions that can be measured to help you determine the short- and long-term results of your program.

Behaviors related to *sexual intercourse* include:

- whether young people have ever had intercourse;
- the age at first intercourse;
- frequency of intercourse; and
- the number of sexual partners in some defined time period.

The box on the next page shows how you might phrase questions to learn about these behaviors.

Contraceptive use, too, has several dimensions that you can measure, including:

- how regularly young people attend clinic appointments (this is called "clinic continuation"), if using a method that requires such visits;
- whether teens continue to use the same method of protection (this is called "method continuation");

- contraceptive continuation, regardless of the method used;
- use of contraception at first intercourse or at last intercourse;
- frequency of sex without contraceptive use; and
- use effectiveness, or whether any method is used correctly.

The box on the next page shows sample questions you could ask to learn about these behaviors.

Measuring antecedent variables

Changes in *knowledge* are easy to achieve and can often be produced rather quickly. It is important to remember, however, that knowledge change is not behavior change. The proof of this is our own bad behaviors. How many of us smoke or do not wear seat belts? Do we know better?

There are plenty of available questionnaires—or "scales"—that measure knowledge of physiology, contraceptives, sexuality, conception, and other topics relevant to teen pregnancy prevention. It is best to secure a scale that already exists because they are more difficult to construct than may be apparent. If you can find one that others

Some questions to measure sexual activity in teen pregnancy prevention programs

Have you ever had sexual intercourse?
1 yes
2 no
If no, stop here.

How old were you the first time?
_____ years old

OR

In what month and year did you have intercourse the first time?
_____/_____
month year

In the last 30 days, how many times have you had sexual intercourse?
_____times

During the last year, with how many different people have you had intercourse?
_____sexual partners

Some questions to measure contraceptive-related behavior in teen pregnancy prevention programs

Clinic continuation, method continuation, and contraceptive continuation are often measured by clinic or medical records, but you can also ask:

How many different places have you been to for contraceptive services in the past year?
_____places

How many different methods of contraception have you used in the past year?
_____methods: Check all you have used:

☐ condom ☐ pills ☐ Norplant ☐ Depo-Provera or a shot
☐ diaphragm ☐ foam or jelly ☐ something else:_____

How many clinic appointments for contraceptive services did you have last year? _____ How many did you go to? _____

When you have sexual intercourse, have you ever used anything or done anything to keep from getting pregnancy or causing a pregnancy?
1 yes
2 no

Did you use or do anything to prevent pregnancy the first time you had intercourse?
1 yes
2 no

Did you use or do anything to prevent pregnancy the last time you had intercourse?
1 yes
2 no

During the past year when you have had sexual intercourse, how often have you used or done anything to prevent pregnancy?
1 all the time
2 most of the time
3 some of the time
4 none of the time

During the last three months, how many times have you had sexual intercourse without using any kind of contraception or protection?
_____

Effectiveness questions about the use of contraceptive methods need to be specific to each method. For example, questions about the correct use of pills should ask about how often pills are taken and what the client does when a pill is missed.

have used, you will get the benefit of their experience.

The most important factor in selecting a scale is to make sure it adequately measures knowledge of the facts you are trying to convey in your program. Some other general rules for scales to measure knowledge include:

- keep them as short as you can, including only the key facts that you want young people or others to know;
- always include a "do not know" response to each question and tell your clients not to guess;
- try to lessen anxiety among your test takers by telling them not to worry about answers they do not know and assuring them that you will tell them the answers later.

Program staff often believe that changes in *attitudes* are difficult to measure. In fact, attitudes are easy to measure but they are often stubbornly resistant to change. There are many available attitude scales relevant to teen pregnancy prevention, including attitudes toward abstinence, contraceptive use, pregnancy, and pregnancy resolution. Again, find an existing scale if you can.

Here are a few other tips:

- keep the answers symmetrical or all on one dimension (see the table below for an example);
- unlike knowledge scales, do not include a "do not know" response on attitude questions—everyone has attitudes;
- unless you want to allow neutrality, use an even number of responses rather than an odd number to prevent respondents from picking the middle or neutral position.

Creating good attitude questions

Not:	But:
How did you like the workshop? 1 excellent 2 pretty interesting 3 a little boring 4 it was awful	How interesting was the workshop? 1 very interesting 2 pretty interesting 3 not very interesting 4 not interesting at all
The responses to this question change dimensions (excellent, interesting, boring, awful)	*The responses to this question stay on one dimension—the degree to which it was interesting—and responses match the dimension asked about in the question.*

Remember that attitudes are not behavior either. Even questions about intent to behave do not measure behavior.

Teen pregnancy prevention programs often seek to develop various *skills*. Popular ones include decision-making skills, assertiveness skills, and problem-solving skills. There are measuring devices to capture each of these skills but many of them are less than satisfying. Still, because the research in teen pregnancy prevention suggests that a skills-based approach is much more successful than an information-giving approach, you should try to measure these skills more accurately.

Skills measurement is improved when you are clear on exactly what a desired skill looks like when it is mastered. A good skills-building curriculum clearly shows or explains the skill for young people, gives them a chance to practice it, takes measures of how well the skill is demonstrated, and then gives additional opportunities to

FIELD NOTES

Avoid arbitrary numbers in your outcome goals

Some programs write outcome goals that sound like this:

50 percent of the parents in the workshop will get 80 percent of the knowledge items correct.

This goal is specific and it certainly sounds like the program has committed to a clear outcome. However, goals like this are arbitrary. At worst, these numbers are plucked out of the air; at best, they are based on some past group of clients.

The real question is how these parents would have done without the program. You should measure success against that standard. As we discussed above, a true control group answers this question. Its achievements represent what would have happened without your program. Without a control or even a comparison group, it is best to write this goal as:

Parents in the workshop will make a statistically significant improvement in their knowledge scores.

Remember that statistical significance means "greater than would be expected by chance" and that it is affected by sample size. In very small samples, changes in knowledge would have to be very large to be significant. In large samples, very small changes in knowledge will be statistically significant, even if these changes are not practically important. Overall then, this strategy is short of the random assignment ideal, but does allow you to invoke a less arbitrary standard to decide which of your results is meaningful.

practice based on feedback. If all curricula were this clearly constructed, it is likely that our skills measurements would improve.

Some teen pregnancy prevention programs believe that the key to preventing teen pregnancy is changing some *personality characteristic*. Perhaps the most popular one of these is self-esteem, although data to show that this is a cause of any pregnancy-related behavior are almost nonexistent. The difficulty with choosing this as an outcome variable is not in its measurement but in how difficult it is to change. Self-esteem is a relatively stable personality characteristic and often requires long-term therapy to be changed significantly. Still, many scales exist to measure self-esteem, and program staff who are determined to measure this variable will find many choices.

Measuring other youth behaviors

Some programs see teen pregnancy prevention as only one of a number of general youth development goals. They may also seek to improve school attendance and performance, prevent drug use, and prevent delinquent behavior leading to involvement with the juvenile justice system. These behaviors too can be assessed by simple questions to young people. School records can also be used to track attendance, grades, or test scores. When young people enroll in your program, get written permission from their parents to gain access to such records.

The chart opposite lists common questions that can measure other youth behaviors besides sexuality or pregnancy.

Collecting your data

The heart of any evaluation is the data you collect in your process and outcome evaluations. You will use them to see whether your program accomplished what it set out to do and whether your program had any significant or lasting impact on your target population. Both process and outcome evaluations have their own data-gathering tools.

Tools for process evaluation

Most teen pregnancy programs take attendance at their various activities and record individual

encounters with their clients. These records are a primary tool of process evaluation. Following are abbreviated versions of various attendance or contact forms.

The forms you choose to use depend on the flow of clients and the convenience of program staff. For example, if a staff member sees many young people individually in a single day, the form for recording contacts with multiple participants over a single day or week is best. If a staff member runs a group activity for the same 15 young people every Tuesday, the attendance sheet is the best choice.

In addition to attendance information, you will also want a way to gather data on the characteristics of teens participating in your program. Program intake forms, like the one on page 163, can be used to gather such data. This example shows how to gather a minimum set of such data, including gender, race/ethnicity, age, and living situation. Program staff will want these critical measures so that they can see if their program is more effective with one subgroup of

Ways to measure other youth behaviors

On delinquent behavior:
Have you ever (or, in some specified time period):

- taken a gun or knife to school?
- taken something from a store without paying for it?
- done anything violent, such as hitting someone, attacking someone with a weapon, been in a fight?
- used a weapon to take things?
- been picked up or cited by the police?
- been arrested?
- done anything to get in trouble, such as set a fire, stolen things, destroyed property, sold drugs?
- carried a weapon?

On drug use:
Have you ever (or, in some specified time period):

- used marijuana?
- used alcohol?
- smoked cigarettes?
- used hard drugs such as crank, crack, cocaine, or pills?
- sniffed or inhaled anything to get high?

You can also ask about the frequencies of these behaviors.

On school performance and attendance:
Have you ever (or, in some most recent specified time period):

- cut school classes without permission?
- been expelled or suspended?
- received any failing grades in courses?
- received honors or awards at school?

Forms for gathering service and attendance data

Recording multiple contacts with a single participant Name of participant _____			
Date of encounter	Kind of contact 1 in person 2 telephone	Codes for Services Provided (listed separately)	Number of Minutes
____/____/____		____ ____ ____ ____	
____/____/____		____ ____ ____ ____	

Recording contacts with multiple participants over a single day or week Staff name _____ Date ____/____/____			
Name	Kind of contact 1 in person 2 telephone	Codes for Services Provided (listed separately)	Number of Minutes
		____ ____ ____ ____	
		____ ____ ____ ____	

Attendance sheet for repeating groups or other events with same participants			
Name	Date of Event	Date of Event	Date of Event
✓ if present	____/____/____	____/____/____	____/____/____
Allen, Joe			
Brian, Kelley			

Group Sign-in Sheet for Events with Different or Unknown Participants Date of Event ____/____/____ Type of Event _____		
First Name	Last Name	Birth Date
		____/____/____
		____/____/____

young people than another. If you wish to, you can add other measures of interest to the form. For example, programs might wish to obtain a measure of socioeconomic status. Mother's education, eligibility for free or reduced lunch at school, and employment of parents are some of these measures.

Tools for outcome evaluation

Two of the most commonly used instruments for collecting outcome data are:

- *questionnaires*, which have written questions to which those responding give written answers; and

A brief intake form

1. Are you...

1 male

2 female

2. What is your race/ethnicity?

1 African American

2 white or Caucasian

3 Hispanic or Latino

4 Asian

5 Native American

6 Other

_____?

(The categories in Question 2 should be adjusted to be most appropriate for your community.)

3. What is your birth date?

_____/_____/_____
month day year

4. With whom do you live? (check all that apply)

☐ mother

☐ father

☐ stepmother or stepfather

☐ guardian or foster parent

☐ brothers

☐ sisters

☐ grandparent

☐ someone else_____
(specify)

- *interviews*, which have written questions that are read aloud to those responding and to which verbal replies are given.

Questionnaires

Questionnaires can be mailed, administered to a group by reading them aloud, or given out one at a time as clients enroll in your program.

In many programs, the most efficient strategy is to administer questionnaires in groups, provided that you try to minimize anxiety and maximize candid response. Young people must be assured of confidentiality.

Many program staff believe that young people will not be honest about their sexual behaviors. In fact, young people will be quite forthcoming if they know the answers to some important questions. These include:

Who wants to know? This question is about the credibility and trustworthiness of the asker. If young people can trust those gathering data, they will be much more candid.

Why do you want to know? The right answer to this question is that you care about each young person and you want to get to know him or her better as you begin to work together. Tell young people that they are enrolling in a program that will talk about some very adult matters in a mature way and that you hope they will begin today to be candid with you.

What will happen to me if I tell you honestly? Young people need to know that neither their chances for program enrollment, your feelings about them, nor your treatment of them will be affected by their answers to any questions you ask them.

Who else will you tell? Teens care whether their friends, and most especially their parents, find out about their sexual behavior. They are also sometimes worried that their teachers will find out about these behaviors. Gather your data anonymously when you can, but always gather it with assurances of confidentiality. You may want your baseline and follow-up measures to include some kind of ID number so that you can match them for the same person (this does not have to be a young person's name).

It is important to stress here that if a program is trying to reduce teen pregnancy, questions must be asked about pregnancy (at a minimum), as well as the proximal variables discussed earlier—that is, whether the young person has ever had sexual intercourse or used contraception. Of course, if the program serves very young children, then these direct questions about sex may be inappropriate. But, remember, a program cannot claim to have affected teen pregnancy rates—or intercourse or contraceptive use—if questions are not asked about these matters.

FIELD NOTES

Tips for gathering sensitive data from young people

Very soon after enrollment, each young person should complete your program entry or intake assessment. Here is how to make that work:

- assign each student an ID number, which you keep on a master sheet next to his or her name;
- write the ID number on the form, making sure that you hand the form to the person it belongs to;
- assemble all your program enrollees, spreading them apart in a large room so that they will have privacy;
- explain what a survey is about, telling them that this is not like a test in school;
- emphasize confidentiality, explaining that their parents or guardians will never see their surveys;
- tell young people that the survey asks them some questions about private issues and that you hope they will be honest and mature in answering the questions;
- ask young people to guess what the private issues are about; they will guess "sex" and "drugs" so tell them they are correct and that you hope they will be straight with you about these issues;
- read the questions aloud, in case some of the young people have literacy problems;
- get helpers to walk around the room, answering questions when students raise their hands but not letting them ask questions out loud.

Analyzing your findings

You are likely to need some help from a professional evaluator to get your process and outcome data analyzed so that you can draw conclusions about the success of your program. However, the logic that underlies the statistical analysis is relatively simple.

You might want to begin by examining your process evaluation data:

- How many clients of various types did you serve?
- What characteristics did those clients have?
- What kinds and amounts of services or interventions did they receive?

This information can be presented against a backdrop of your intended targets and service delivery strategies. Did you reach those you intended and did you deliver what you planned?

Then, you might turn to your outcome analysis. If you have a control or comparison group, you will want to make sure that it does indeed match the program group as closely as possible:

- Do the two groups include about the same percentage of males and females?
- Are they similar in race/ethnicity?
- Are they of the same average age?

Your evaluator will want to use tests of significance to make sure that any differences are greater than would be expected by chance alone. He or she will use your baseline data to answer these questions so that you are sure that the comparison or control group can be fairly compared with the program group.

With this assurance, you can begin to examine whether the changes in your program group are greater and more positive than the changes in your control group. You may also want to examine whether the positive changes you find in your program group occurred among all subgroups:

- Did the males make as much progress as the females?
- Did the younger teens learn as much as the older ones?

Again, you will need some help from an evaluator to know which particular statistical tests will help answer these questions. The chart on the next page will help you understand the principles behind the statistical analyses your evaluator will conduct.

The essential steps in data analysis

1. Compare those lost to follow-up with those you kept up with for the whole period of the evaluation.

You will lose some clients and you need to find out whether they are different from those from whom you have complete data; use your baseline data to answer this question. Remember that one measure of program success is your retention rate. If you are losing many clients who do not really get the intervention, you need to know that and know who they are.

WHAT YOU HOPE TO FIND: Few losses and no apparent difference between those who stay and those who drop out.

2. Compare your clients with those in the comparison group.

Use baseline data again to see if you really got a good comparison. Are your clients the same as the comparison or control group? Were they the same when you first measured them?

WHAT YOU HOPE TO FIND: No difference

3. Compare changes between your baseline and post-program measures of outcomes and impacts for both your program clients and their comparisons.

Did your program group change more than your comparison group? We hope so. It is not enough for them to be different in the end. They must have different rates of change. Sometimes when baseline is not appropriate or possible then only post-program measures are enough (like school dropout rates).

WHAT YOU HOPE TO FIND: More changes in your program group than among comparisons.

4. Introduce necessary controls and do subgroup analysis.

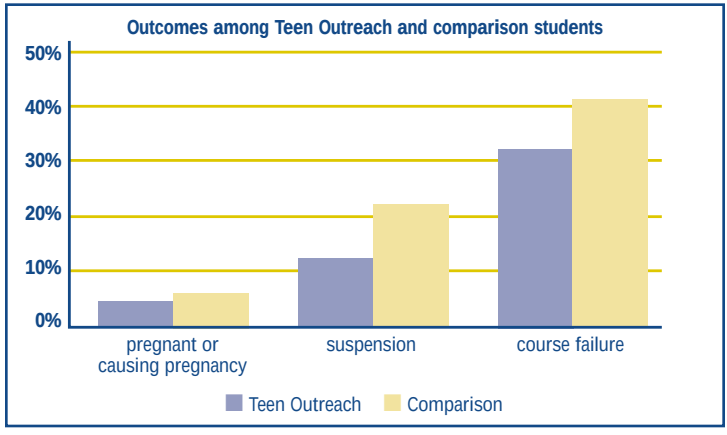
Here you can begin to ask things such as whether your findings hold for all clients and whether these findings persist when you introduce controls for baseline or other differences between the comparison group and your program group.

WHAT YOU HOPE TO FIND: The advantages of your program group persist in all subgroups and even after controls.

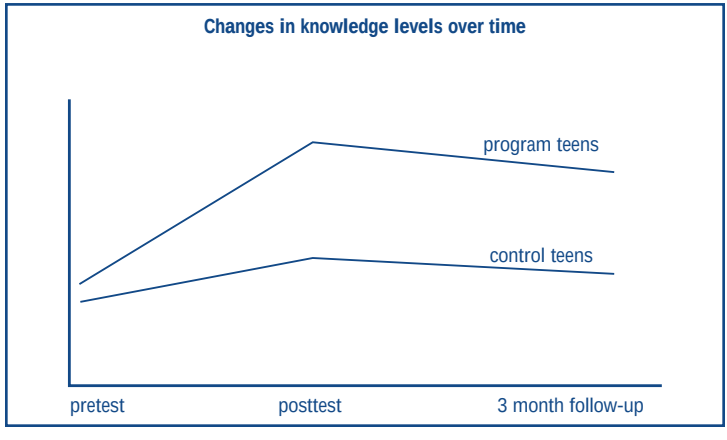
Conclusion: Presenting your findings

After completing the data analysis, you and your evaluator will want to present your results in a simple and compelling format. Because many people have low tolerance for lots of tables and numbers, it is sometimes wise to use such devices as bar graphs, line charts, or other devices for showing your results.

The chart below shows some outcome results from the Teen Outreach Program (Philliber & Allen, 1992). Notice how easy it is to understand these results.



The following graph illustrates another simple way to present outcome data, this time from a knowledge change program. Notice how easily you can see that the knowledge of those in



the program group went up over time while the knowledge of the control group did not.

Your evaluation report is likely to have many different audiences, so remember that it should be written (either by you or your evaluator) so that those without technical training can understand and learn from your findings. It is possible to learn from both positive and negative findings, so the report should present both. If findings are positive there is always rejoicing and funders may have renewed confidence in your work. If the findings are negative, present your hypotheses about why this happened and outline a plan to change the program for the better. Even these negative findings, as embarrassing as they may be for you, can help build a better science base for how to prevent teen pregnancy.

References

Brown, S.S., & Eisenberg, L. (Eds.). (1995). *The best intentions: Unintended pregnancy and the well-being of children and families*. Washington, DC: National Academy Press.

Card, J.J. (Ed.). (1993). *Handbook of adolescent sexuality and pregnancy research and evaluation instruments*. Newbury Park, CA: Sage Publications.

Frost, J., & Forrest, J.D. (1995). Understanding the impact of effective teenage pregnancy prevention programs. *Family Planning Perspectives*, *27*:188-195.

Kirby, D. (1997). *No easy answers: Research findings on programs to reduce teen pregnancy.* Washington, DC: National Campaign to Prevent Teen Pregnancy.

Moore, K., Sugland, B., Blumenthal, C., Glei, D., & Snyder, N. (1995). *Adolescent pregnancy prevention programs: Interventions and evaluations*. Washington, DC: Child Trends.

National Center for Health Statistics. (May 1997). *Fertility, family planning and women's health: New data from the 1995 National Survey of Family Growth.* Washington, DC: Author.

Philliber, S., & Allen, J. (1992). Life options and community service: Teen Outreach Program. In Miller, B., Card, J.J., Paikoff, R., & Peterson J. (Eds.). *Preventing adolescent pregnancy.* Newbury Park, CA: Sage Publications.