Planning Your Local Evaluation
May 24, 2012
3:00-5:00pm EST
Transcript

LINDSAY: Good afternoon. My name is Lindsay and I will be your conference operator today. At this time, I would like to welcome everyone to the Planning Your Local Evaluation webinar. All lines have been placed on mute to prevent any background noise. After the speaker's remarks, there will be a question and answer session. If you would like to ask a question during this time, simply press star, then the number one on your telephone keypad. If you would like to withdraw your question, press the pound key. Thank you. Dr. Olivia Ashley with RTI International, you may begin your conference.

DR. OLIVIA ASHLEY: Thank you, Lindsay. Welcome, everyone. I'm Olivia Ashley from RTI International in North Carolina. And our webinar today is for State and Tribal PREP grantees about planning your local evaluation. My co-presenter today is Marnie Kahn from RTI. And our co-authors at RTI are Ina Wallace, Lizette Savadre, Laurie Palin and Kathy Williams. But today, Marnie and I are going to be doing all of the talking about the content that our coworkers helped to develop. And you can see that Marnie is your contact person if you have any questions or want any follow-up after today.

All right. So I wanted to share with you the last count that we've seen. We have 141 folks registered for this webinar. We perused the spreadsheet of all the folks who signed up. And it looked, at last count, that we had about nineteen evaluators and a whole lot of program administrators and program staff. So today our primary audience is program administrators and program staff.

I'm seeing from the registration list and from the participant list, we have some very highly respected, knowledgeable evaluators on today's call. And so for several folks on the call today, it may just be a refresher. And it will help to make sure that everyone is on the same page, the program staff, the program administrators and evaluators.

We are happy to take questions or comments at any level. So if you have a complex question, please don't hold back. And if you have a simple question, please don't hold back. Because lots of different people on today's call are going to have lots of different levels of experience and different perspectives and different backgrounds.

We have four main topics we're going to talk about today and those are evaluation design samples, measurements and institutional review board or IRB issues. Our goal for today is for program staff and evaluators who are on the call to be able to weigh pros and cons of different outcome evaluation designs. And for everyone to have a good understanding of what to consider when thinking about defining your evaluation sample. And that is different from your intervention sample. And so Marnie will be talking about that piece.

We also want you to be able to ensure that the measures that you're using for your PREP project are appropriate and also be able to figure out whether IRB involvement and review are recommended.

So here's the game plan for what we're going to do today. You can see the different estimates that we've given about how much time we're going to spend. If we get tons and tons of comments and questions, we may have to cut some of that off. But we can keep a record in the chat box that you see on the bottom right of all the remaining comments and questions that you submit and we may be able to work with FYSB who is also on today's call for some possible follow-up like some frequently asked questions and answers or some other way to make sure that we address folks' needs.

We're going to try to stick to these timelines today. If we actually have 141 people on and everybody has a question, we may not get to them all, but we really want to hear what questions and comments are. So we're going to allow time periodically for questions and comments. Our first line of defense is going to be that chat box in the bottom right.

So as we are talking today and questions or comments are coming up, we would really encourage you to type in your question or your comment there. And when we pause for questions and comments, Marnie and I will be furiously reading all of your questions and comments and seeing which ones would be good ones for us to try to address on the call. And if we see no comments or questions, then we're going to ask Lindsay to open it up for live questions and folks will be able to press a button on your phone to get Lindsay's attention and then she'll let you just say your question out loud or your comment out loud instead of typing. So we'll start with chat box. Feel free to type anything that you like in the chat box as we go along.

The first thing we just want to make sure we're all on the same page about what we're talking about today. Today our focus is on local evaluations, which is your project's specific assessment of your interventions or curricula or activities. And that is different from the PREP multi-component evaluation which is a national evaluation across grantees.

We are talking about what you and your local evaluator are working on together just for your PREP project. And this could include implementation or process evaluation. And the main purpose of that is just to describe your project. Primarily, so that if you find some great effects, you and others have documentation about how to replicate that and everybody understands exactly what you did when you got your great results. So they can do that too.

We're also going to talk about outcome evaluation. And that's where you are trying to identify whether your program is effective in preventing teen pregnancy.

There may be lots of reasons that you're conducting your local evaluation. Certainly, it's important to learn the extent to which your project has achieved its objective. The evaluation results can tell you how to improve your project. Certainly, you are accountable to FYSB and other stakeholders in your State or Tribal community. And the evaluation results could provide some evidence or data or information that would support future funding for your project. And FYSB really sees the state and Tribal PREP project a golden opportunity for you and them to contribute to the evidence base for teen pregnancy prevention.

You're doing some very innovative things out there that are very tailored to your communities. And I think many of you know that many of your communities are not well represented in the literature. It's a real opportunity for diverse populations to have a voice in the literature about what works and how things work.

So some key considerations for evaluation planning. We've boiled this down to two. There are a million considerations, but we don't want to overwhelm folks. And we want to try to keep this to two hours today. So one thing is, who's going to work on your evaluations? And another is, what do you want to learn about your project? Much like you're tailoring your project activities, you're probably going to want to tailor your evaluation activities based on what you think is important to learn.

And FYSB wants you to also consider, who is the audience for your findings? That's really important. If your audience is your community, then the values and priorities in your community are something that you're really going to want to address. If you're also thinking about a larger audience like the field or peer-reviewed manuscripts, there may be additional things that you want to learn. And there are certainly lots of other audiences in between.

Many of you have already planned your project and your evaluation. And so this slide is for those who have not. We would like for you to keep in mind that projects that are planned in tandem with local evaluations leave open many more options for rigorous evaluation designs. Then when the project is planned and the evaluation comes in as an afterthought or the evaluation is planned first and then the project is put on top of the evaluation.

So we don't think here about project planning and evaluation planning like that cartoon at the bottom on the left, where one comes before the other and you can see how confused they are. That is not a good model.

On the right hand side, we think more like the stick people who are super happy that they planned their project and their evaluation together. So we really are thinking more of a collaborative approach where the planning for both is occurring at the same time and the project needs are getting met and the evaluation needs are getting met. There may need to be some compromise, but it's not one is more important than the other.

So this slide has some data from the needs assessment that we did in February. Almost all PREP grantees contributed to that needs assessment. So we really want to thank folks for responding to the survey needs assessment that went out. And because this is in February, things may have changed since then. So we know that the [numbers] may look very different.

But on the left-hand side, you can see in February about 85 percent of grantees said that they plan to conduct a local evaluation. And then the pod on the right is of those 85 percent, you can see that 57 percent of those folks had hired a local evaluator, but 43 percent had not yet hired a local evaluator.

And different types of PREP grantees are at different stages. Some folks were awarded grants earlier than others. And they're just maybe at different places in terms of planning and hiring. So this actually made a lot of sense to us in terms of where folks were.

This slide right here references a handout that was sent along with the webinar slide. And this is a handout on suggestions for identifying an appropriate local evaluator. I just wanted to mention that this handout came along and highlight a few of the TIPS that we included on this handout.

The second, third and fourth bullet are all on this slide about places where you can find a local evaluator if you are still shopping or if you have some fact turnover issue that occurs later on in your grant.

We certainly recommend contacting evaluators or researchers who've published evaluations of adolescent reproductive health programs, or professional associations. Because lots of professional associations have listservs. And the handout has lots of ideas about professional associations that you may want to think about looking at their website.

The National Institutes of Health Website has a search function where you can look there. And we've actually heard that NIH project officers, I think at NICHD, are reaching out to some of their current or former grantees who may have done evaluation, letting them know that because funding is very tight at NIH, they may want to be aware of opportunities like teen pregnancy prevention. They actually mentioned that recently to NIH researchers.

So NIH researchers who usually get funding through NIH may be less and less able to use those government funds. They maybe more and more available for you to reach out to as potential evaluators.

And then some tips at the bottom of the slide, one of the things that we recommend is asking a potential evaluator about how successful they've been in the past recruiting comparison groups. Because this can be very challenging even for an experienced evaluator.

Also, it's important to talk about what your timeline is and how much time you expect this person to spend on your project and what products you are looking for. Are you wanting a report? Or are you wanting a presentation to your community? And what would that look like? So I think it's pretty important at the beginning to talk about what your vision is at the end of the project. And we always suggest that you ask for references from other program staff to see what their experience was like working with that evaluator.

All right. So two quick take home points so far. We recommend planning the project and the evaluation together. And we recommend making sure that the evaluator that you hire and that you’re working with meets your needs.

So at this point, we’re looking in the chat box to see if there are any comments from folks or any questions about anything that we’ve talked about so far. We’re not seeing any chat in there. And I’m wondering, Megan, have you seen any there that we don’t have?

MEGAN: No, no questions.

DR. OLIVIA ASHLEY: Lindsay, would you like to open it up to see if there are any live questions or comments before we move on?

LINDSAY: At this time, I would like to remind everyone in order to ask a question, press star, then the number one on your telephone keypad. There are no further questions at this time.

DR. OLIVIA ASHLEY: Sounds great. All right. Well, let’s move on and talk about the first substantive topic which is outcome evaluation designs. So our goal today is not to overwhelm program administrators or staff with lots of detailed information about conducting evaluations that the local evaluator probably should know and probably already knows, but just to make sure that program staff have enough information to be comfortable communicating with the evaluator and asking the right question.

While there are many considerations in selecting an outcome evaluation design, we’ve boiled our list down to two basic questions, but you may be thinking of many other things. One is ... and this is about comparison groups ... are you able to delay project activities for some teens? Or are you able to withhold project activities for some teens?

Some communities may be willing to let you do that, but not everybody gets the program at the same time. Or maybe not everybody gets the program. And some communities may feel very strongly that everybody needs to get the benefit of this as quickly as possible.

The second consideration is if you are able to delay or withhold some project activities from some teens, can you get cooperation from your community partners to randomly assign teams or schools or groups to participate in project activities or service control?

And so when we talk about randomly assigned, we simply mean that your partners let you decide who gets what when instead of them deciding. It's not that if you have a set of schools, these schools are saying, "Well, we've got to get it right now," and these other schools are saying, "We don't want it right now." It's more that you say, "Will you let me decide who gets the program now and who doesn't?" And if they agree to that, then you're able to do random assignments.

We always recommend that you plan for the most rigorous design that's possible. Because I'm sure everyone knows from working on this project and other project that sometimes things just don't work out the way you want. And so you may need to go to a Plan B or a Plan C. But if Plan A was not rigorous to start with, you may be stuck with something on your Plan B or Plan C that you really don't like. So we recommend starting as strong as possible. And then if that plan doesn't work out for you when you fall back to a Plan B, it may still be moderately strong.

So we think that you should start at the top and know that that may not work out for you and you may end up with a perfectly good evaluation design. But start as strong as you possibly can. And there are lots and lots of different evaluation designs. We're focusing on three main categories and they are experimental or randomized control trials, quasi-experimental. And this is when you have a comparison group, but it's not randomized. And pre-experimental where you don't have a comparison group. And so we'll talk about these three.

We're starting at the bottom which is not where we recommend that you start. But we just want to compare and contrast if you start at the bottom, what would some of the issues be? And then, of course, if you had to go to Plan B, you might have even more issues.

So this first illustration is two designs that are not rigorous. They do not involve a comparison group. So you can see the purple box, which is your project, delivers the program to everybody. And then at the top, you collect data at the end or you collect data at the beginning. And then you deliver your project to everybody in the purple box and then you collect data again.

So this bottom one is a pre-/post-design with no comparison group. And the considerations for this are that it's very difficult to say that participation in the project was associated with the change in outcomes. If you saw improvements in teen pregnancy or if you saw things getting worse among your target population, it's difficult to know whether it's your project that was the cause of that because you don't have a comparison group to see what would have happened if your project was not there.

So what's happening here is, the two yellow boxes are just measuring changes over time where you're collecting data. You're doing something and you're collecting more data. But everybody is getting something. So you don't really know if they didn't get the something what would have happened.

So here we don't know if the outcome is because of the project or maybe because of something else. I think folks know that teen pregnancy rates are dropping in the nation overall. And so maybe the rates in your community are following a national trend and it's not about your project. Or maybe they're dropping, but they would not have dropped quite as much if it wasn't for your project and you have no way of knowing that.

The second design is more rigorous and this is a quasi-experimental design. And this you can see you have a project box and then you have a comparison group. So this looks really similar with the addition of a comparison group.

The first thing you do is collect some data. And then some people get the project and some people do not get the project or they do not get it right away. And then you collect your data. And after evaluation ends, you can do anything you want with that comparison group. You can deliver your project activities six months later if that's when your evaluation data collection is ending or one year later or two years later. Whenever it is that you've decided that you're finished collecting data, you're free to deliver project activities to those folks.

So in this case, the units ... and units is not a very personal term, but we mean teens or schools ... are matched in the project group with those in the comparison group on key variables. And they may be demographic variables. Anything that you have that can show you some data that says that these two groups would be similar is what you're trying to do is to get them matched.

So what you need at the bottom here is you need to have a surplus of teens or schools for your evaluation. You've got to have some folks that you can either not deliver a program to right now or not deliver a program to. Not that everybody gets the same thing at the same time. And you certainly have to coordinate this data collection with the program implementation. You didn't see, for instance, project and then the baseline data collections and then the follow-up data collection. You've got to be able to collect your baseline data first and then have some people get the program and not get the program.

So I mentioned matching. The question here is what if the comparison group is not matched? And we've put down some examples of some non-equivalent comparison groups. Certainly, you know your communities very, very well. You've done needs assessments. You can look at different neighborhoods or communities or community based organizations or schools. And the folks that I met with at the conference can tell me immediately if the schools that they have on their list, which schools were similar and which schools were very, very different on different characteristics.

So those would be an example of a non-equivalent comparison group. If you have schools that you know the kids in the two different schools are very, very different from each other. But there are some other non-equivalent comparison groups that we really want to call your attention to.

One is teens who may have refused to participate in your project. Probably not a good idea to use for your comparison group. Because these are teens who for one reason or another are not motivated. Maybe logistically they were not able to participate in the program. But there's something different about them than the students who completed your program or were willing to at least start it.

And then I mentioned also completing the program. Teens how drop out of your project activities are also not comparable to the kids who stayed in your program.

Certainly, you can use county, state or national data. But I think many folks on the call will acknowledge that the communities that you are targeting may not look like the county or the state or the nation. And in fact, you may have picked your target population for very specific reasons, very specific needs or values that they share with the spirit of the grant that you're doing.

So in terms of a matched comparison group, the larger county, state and national data are probably not matched. And historical data may be available. It may be that there are some survey data that are available for the schools that you are going to be considering in your evaluations. And maybe they completed some information that looks like what you want a year ago or two years ago. And you'd like to just compare your teens in your projects to what those scores look like a year or two ago.

As we all know, things change over time. I mentioned that teen pregnancy rates may have fluctuated during that time. Or there may be seminal events. I remember when we were working on a teen pregnancy prevention project, Britney Spears' younger sister got pregnant and it was all over the news. And attitudes shifted widely about what teens began to think about teen pregnancy. Because this was someone who was on Nickelodeon and very well respected. And it was a very different feeling at the time. So if Jamie Lynn Spears shows up between your historical data and your data that you're collecting now, that may be what's happening or at least contributing to it, not just your project.

So in all these cases I've mentioned, if you see differences with a non-equivalent comparison group, it's hard to know if what you're seeing is because of your project or something else.

And so, we recommend that you work with your evaluator. After you've done the best matching that you can and you've collected your baseline data, to test for baseline differences between your teens who receive your project and your teens who did not receive your project. And see if those groups really are the same. If they're not the same, figure out how they're not the same. They may be different in terms of race/ethnicity or age or parent involvement or some data that you may be collecting at the beginning about sexual activity. And you may need for your evaluator to control for those statistically when you analyze your data.

The last design we've got here is an experimental design or randomized control trial. And you can see it looks almost identical to the plain old comparison group design. The only difference is that green R that you can see that comes after you've collected your baseline data.

What you've got there after you've collected your baseline data is that you are randomly assigning who goes to your comparison group and who's going to receive your project activities now. That's really the only difference here. Instead of having a comparison group who decides for themselves whether they're going to get the program now or later, you get to make that decision.

And you make that decision randomly by tossing a coin or saying for each teen that we enroll, the first teen is going to go into getting our project and the second teen is going to wait. And the third teen is going to get our project and the fourth teen is going to wait. So that you are making some decisions very arbitrarily. It's not based on any characteristic of the teens themselves.

And you can see at the bottom the same exact requirements. You still need to have more units, teens, schools, clinics, community based organizations than are going to receive the project all at once. And you have to coordinate with project implementation.

So, no big differences here between a randomized control trial and a comparison, just a plain old comparison group design except that you're making decisions about who gets what.

Several folks may know that the randomized control trial is considered the gold standard. And so we thought that we'd try to explain why some people consider it to be the gold standard. It's definitely the best method to control for differences between groups before you start delivering your program.

And some of those differences we mentioned, you can measure. You can measure kids report to their race/ethnicity and report to their parent involvement, what kind of grades they're making in school, that type of thing. But there may be other differences that you cannot measure.

There may be characteristics of schools, how much funding the school is getting, characteristics of kids motivation to be in your program, other things that you may have never have thought of. And so if you're randomly assigning kids or schools or groups, then you don't have to think about, well, what is the difference between this group who wants program now and the group who doesn't want the now? It's being randomly assigned.

This supplies the best report for ruling out any other explanations for outcomes that are not related to your project. And it makes your results much more convincing and understandable to the field and policymakers who are trying to decide whether your evaluation results are something that need to be replicated elsewhere.

Some other design considerations. And these maybe things to talk about with your evaluator. It's really important to attend to and try to prevent contamination between your two groups. And contamination means that if you have a comparison group that they're not somehow getting the effect of your program.

It could be that you want to make sure they don't accidentally end up in classroom sessions. That would be the best way they could get your program. But it could also mean, for instance, if you are thinking about delivering the program to some kids who are taking health class and some students in the same school who are not taking health class right now.

Those kids may be dating each other, certainly talking among each other. They may be best friends. They may be making decisions about whether to be sexually active with each other. And so if you are kids who are receiving your project activities are being protected and they are starting to stop having sex or refuse or say no or begin to use contraception, they are going to make your comparison group start to look more like them. And it makes it very difficult to see whether you have a difference between your comparison groups. It may be great for the community, but it's very difficult for your evaluation results to show that it was because of your project activities that the treatment group or the kids who got your project activities got better.

It's also important to think about matching, I mentioned, of the groups. And then we have this term "differential attrition." And that just means that teens may drop out of your evaluation at different or among different subgroups. So you may have done the best job in the world at matching schools or communities and then you go out to try to get teens and their parents to consent to be in your evaluation. And you get them in your evaluation and everything is looking great at baseline. And then some teens start to drop out of your evaluation. It happens. They move. They decide they don't want to complete your evaluation forms later on.

And when that happens, it may be that the two groups start to look different. If that is the case, that is another analysis that your evaluator can do. It's called an attrition analysis. And so it may be important to look at characteristics of teens that are associated with who dropped out and to control for those. And that's something your evaluator should be able to help you with.

So we also think it's really important to talk about the evaluation design that you are aiming for and why that evaluation design is important with your subawardees and organizations who are partnering with you. And regardless of the design or the end users of your findings, it's important to just understand and acknowledge the limitations of your design.

Every design has limitations. There is no perfect evaluation design out there. And we all know that these are real world studies. These are not laboratory experiments. And so if there is something that is not going an ideal way, that's okay. That's part of why you started with the most rigorous design that you could. And it's just important to be aware of that and document what the limitations are.

All right. So we've got a quick quiz. This is a poll. And so we would like for you to answer what you think in terms of this multiple choice question. Everybody should be able to see a poll over on your right, which is the exact same question over here on the right. And I'll just read these out. And as you decide which of these answers you like the best, we'd like for you to click on the answer. And we're seeing some people are answering which is fantastic. So choice A, improvements among teens can be attributed to project activities even if there's no comparison group. B, if there's no comparison group, a pretest, post-test design simply measures changes over time. And C would be that both A and B are true.

So we'd like to see what folks think. And the poll is timed. And it looks like we have a few seconds left. So we'll let it finish itself out and then we'll see what people came up with and talk about answers.

Okay. The poll closed itself. So of the folks who responded, the poll is now calculating for us what the response looks like. It looks like it's going to take about eight seconds for you to see that on the screen.

Okay. So I hope everybody can see the results of the poll. It looks like we got a small number of responses to Point A. Improvements among teens can be attributed to project activities even if there's no comparison group. So we're really glad that very few people chose that one because that is not true.

The second one, B, if there is no comparison group, a pre-test, post-test deign simply measures changed over time. And that is true. What I was saying before is that if you collect data, deliver your project to everyone and then collect data again, what you're measuring is what happened over time, not what happened in comparison to people who did not get your project.

So it looks like a lot of people, about a third of folks who are on the call, said that B was true. And then it looks like about a quarter of folks put down that they thought that maybe both A and B were true. So we just want to reinforce that you cannot say that it was your project's responsibility for what happened with teens if you don't have a comparison group. And that is what A says. A is not true.

And then we have a lot of folks who were not able to get their answers in before the poll closed on itself. So I want to just ask one more question. And why don't we wait to open the poll since it closes itself? And give folks a chance to think about it for a second and then we will open it up.

So which of the following is required to successfully implement a randomized control trial? A is that more schools or teens are needed than for a quasi-experimental, non-randomized comparison group design. B is that a larger budget is needed then for a quasi-experimental design. C is that schools or parents must agree to random assignments. And D is all of the above.

All right. So we're opening the polls. And we'd like for you to put your best guess for what you think there. You know, when we practiced, these polls gave us five minutes. Now that we're online, they're giving us one minute. So we're glad we gave you a chance to read it and think about it before we set that one minute timer. And we get to watch the bar graph here. And we're seeing the no answer bar graph go down and the other bar graph go up which is great. We've got about six seconds left on the poll. So get your answer in and then the computer will calculate up.

All right. The poll is closed and now it's calculating up so you can see the bar graph. Everybody decided. All right. So here are the results that we have here. Which of the following is required to successfully implement a randomized control trial? So it looks like we have about seven folks who thought that more schools or teens are needed than for a quasi-experimental or non-randomized comparison group design. And I've heard that before, but that is actually not true.

If you remember on that slide, the only thing that was different was the R in those models that we had. So you don't need more teens or more schools if you want to randomize, that is not a requirement. So for folks who might have thought that that was a barrier to doing random assignments, that may open up some possibilities for you.

The second one, nobody fell for the larger budget. Fake answer. So, that's good. Larger budget is not needed. It's the exact same thing, but the answer is C which is schools or parents have to agree to random assignments. And so in some cases if you are randomly assigning schools, you're going to need to work with a principal or an administrator or the district or someone who will allow you to decide which schools get what and when.

If you are doing individual assignment, let's say you work in a clinic or a community-based organization where you don't have contamination problems. All the teens are dating each other. They may just be in waiting rooms or after school. But then they disperse in their free time. Then you may need to have parents who are providing consent for teens in your evaluation to let you decide who gets project activities now and who may get them later.

And so there were maybe half of folks who answered all of the above. So it looks like we may have been able to trick a few folks. And we were not trying to trick you, but we had heard some of these myths before. And so we just wanted to try to address some of those.

So the answer is C that the schools are going to have to agree to random assignments. And if you have an active consent procedure, then the parents are going to have to agree to random assignments.

All right. Thank you guys very much for answering the poll. It's really helpful to see what your thoughts are about this and we'll take some comments and some questions in just a minute.

So here is our summary for this section.

Start with the most rigorous design that's possible. You cannot attribute changes over time to a project without a comparison group. The only difference between a plain old comparison group design and a randomized design is that you decide who gets the project activities or when they receive them. Try to make sure that your comparison group doesn't get contaminated by your treatment group. And remember that no evaluation is perfect. Don't get discouraged or overwhelmed if something happens or you see a limitation that you wish you could get rid of. Just be aware of what those things are and document that.

All right. So we're looking at the chat questions and comments. And so I see a couple here. One says, "I know in some communities the number of teens is very low like 100 or less. Would you still suggest a comparison group be used?" I'll throw out my answer and I'll let Marnie see if she has anything to add. FYSB may want to weigh in on this. Or there may be other folks who want to chat and put their thoughts in because there are lots of folks on the phone who have a lot of experience. Some may have more experience than we have.

My thinking of that is that if you have 100 or less teens, it may be very difficult to do an evaluation where you have statistically significant differences if you're dividing that 100 into an intervention group and a comparison group. But one of the things you may want to think about is finding a comparison group that is outside of your community, but that is similar. And I know that that's challenging. There are not two communities that are exactly alike on the planet. But if you find a community that is similar, I know there are lots of communities that are matched with sister communities who are very similar to them. I could easily tell you if I was trying to match where I live with somewhere else, I could think of some places that feel like where I live and could be a possibility. And then if you get something like 200 teens, you actually may have enough sample. And Marnie’s actually going to talk about sample in just a minute. So that’s my thinking on that. Marnie, did you want to add or disagree or anything like that?

MARNIE: I don't disagree. The only thing I would add, which we’re going to talk about in a few minutes, is that you can figure out whether 100 teens would be enough or not by conducting a power analysis and working with your evaluator and perhaps a statistician. We’re going to talk about that in a minute. But it’s definitely difficult with that sample size. But there have been studies that have that many teens. And there may be particular measures or outcomes that you would be able to look at.

And if you had a design like Olivia was talking about where your control group or comparison group within that community of 100 teens still receive services in a delayed fashion, you wouldn’t have to withhold anything from your community and you could still involve everybody. And you could still collect lots of interesting information on the full community, even if you don’t have enough power which is what we’re going to talk about in a few minutes to look at some of the outcomes that you’re interested in within the 100 teens. But I agree that looking for a comparison community potentially outside of that might be a good approach as well.

DR. OLIVIA ASHLEY: Okay. And I see a second question, and that is ... and this is even trickier. “What type of evaluation is required for PREP awardees?” So I’ll share with you what I’ve heard. But you know who really need to say if there is a definitive answer here and that’s probably FYSB. What I’ve heard is that FYSB is very eager to seek contributions to the field and to the evidence base.

And so to make a significant contribution to the field, we would recommend ... this is not a requirement ... that you start as rigorously as possible. But I’ve not heard any specific requirement in terms of everyone must do their evaluation a certain way because all the grantees and their communities and their projects are very unique and innovative. And so I’ve not seen any restrictions or limitations or requirements that everybody do things a certain way. But I’m wondering, LeBretia or anyone else from FYSB if you’d like to comment on that.

SARAH: Hi, this is Sarah from FYSB. The only requirement in terms of evaluation or from an evaluation perspective I guess I would say is that PREP grantees complete the performance measures requirements that have been outlined in the performance measures documents that you've all received information about. There are not separate requirements in terms of evaluation or in terms of a local evaluation for PREP grantees.

DR. OLIVIA ASHLEY: Excellent. Thank you, Sarah.

SARAH: You're welcome.

DR. OLIVIA ASHLEY: And the questions are starting to come in which is great. Here's one. "If an evidence-based program already is proven to work, what is the use of doing more research-oriented evaluation effort instead of investing all of the money on implementation with fidelity?" I actually heard this discussion at the conference among lots of different folks. I'll tell you my thoughts on that. And other people may have lots of different thoughts on that.

An evidence-based program may have been proven to work in a specific community at a specific time under specific conditions. For example, we've worked on lots of evidence-based programs that were delivered before No Child Left Behind. And if they were delivered in schools and No Child Left Behind hit the nation, wow. Things really changed in classrooms. And so that type of a change in terms of restrictions on what schools are able to do can make a huge difference in how you're able to deliver your program.

Also, if you are delivering your program to a different population, then the program was originally evaluated and delivered to. Or if you were doing something innovative like incorporating adulthood preparation subjects, you may see a different type of response to the program. It may be an even better response, but there can be lots of reasons to look at evaluations, to contribute to the literature. There are so many questions that are not answered. If we had them all answered, I think we wouldn't have any teen pregnancy. So it feels like there are lots of reasons.

I also totally understand that many service delivery staff folks are very much prioritizing implementation with fidelity instead of evaluation. So I see both sides of that situation. Marnie, do you want to add to that?

MARNIE: I just pulled up the slides that Olivia covered in the beginning about reasons to conduct the local evaluation. And some of them may not be necessarily just about adding to the already existing evidence on particular curricula. Your community and stakeholders might want information about how the program is working in your community. It might help you get future funding as Olivia was talking about earlier. And it can help you make improvements to your project along the way. So there are lots of reasons to still conduct evaluation of some sort of you're using an evidence-based curriculum.

DR. OLIVIA ASHLEY: So it feels like there are a lot of questions out here like this. Because the second question underneath that one is the same exact thought. So we definitely see folks who have dialogue about this. We want to move onto the next question we've got to measure differential attrition. What we need to track each teen specifically. And if so, how have you done this successfully without violating teens' privacy and IRB?

So I can tell you what our experience has been. And then we can talk for two hours. This is one of our favorite subjects is tracking and attrition. Marnie and I are just big evaluation nerds and that's all we like to talk about in our spare time, which is very sad.

But I can tell you what we have seen. And everybody may do this in a different way and your local evaluator may have different ideas. And your community and your IRB may have different constraints. What we have done is we have signed an ID to every student or teen who is in our evaluation. And that ID number appears on their questionnaire, not their name or their initials or their date of birth. And everyone who is enrolled into our evaluation whether they're a comparison group or intervention group fills out the baseline questionnaire.

And then half of the teens go and get the programs. And half of the teens either don't get it right now or maybe they get something else that's not related to teen pregnancy prevention. And then at some later point, we try to survey teens again. And at that point, we do have a roster that has the teens' name and their ID number. We've put that under lock and key so that nobody can see that except for us. And when we pass out the questionnaires, we hand them their questionnaire that has their I don't number on it. Some teens are no longer in school or don't show up or may refuse. Because our IRB says that everything needs to be voluntary and teens can refuse to answer any question or the whole questionnaire if that's what their choice is.

And so we are measuring attrition by looking at the baseline data of the kids who were at baseline, but who were not at follow-up. And we look at the characteristics of those kids as a group versus everybody else who did show up for follow-up as a group. So we are not saying, "Oh, well, John Doe was Chinese and he was seventeen years old."

We're not looking at that. We're looking at the group, at their demographic characteristics and seeing whether the groups who dropped out is different from the group who stayed in. And when we say differential attrition, we do that among the treatment group and we do that among the comparison group as well. So that's how we've done it without violating privacy or IRB. Marnie, did you want to add anything to that?

MARNIE: The only thing that I would add is that differential attrition can also just include how much dropout there is. So sometimes receiving project activities is a great way to keep kids in your evaluation. And so you have less dropout among teens who receive project activities than teens in your comparison or control group.

Sometimes it's the other way around. The teens who are being part of the project all the time don't want to do one more thing. And so they don't want to be part of the follow-up survey. And the comparison group teens, that's all they're doing potentially for you if they're not receiving some alternative from you. And maybe they want to be part of the survey. So it's possible that there would be just different rates of dropout in the two groups and that's another thing to look at in addition to their characteristics.

DR. OLIVIA ASHLEY: I'm glad you mentioned that. Because when I have done evaluations for federal funders, the first question they've asked when we got the baseline data is what was the response rate in the treatment group? And what was the response rate in the comparison group? And that's exactly what Marnie's talking about. And then we start looking at the characteristics of teens.

These are really excellent questions and comments. So I'm going to stop talking and let Marnie take over at this point. But please keep your questions and comments coming so that we can have some dialogue this way.

MARNIE: Okay. So next we're going to talk a little bit about defining an evaluation sample. And by sample, we mean the people including teens or perhaps parents or other individuals who are providing data for your evaluation.

The first consideration in defining your evaluation sample that we would like to talk about is distinguishing your evaluation sample from your intervention sample. So your intervention sample includes teens or others who are receiving project activities or services from your PREP project. And this doesn't necessarily need to be the same group of individuals as the folks in your evaluation sample. So your outcome evaluation sample includes anybody whose data are contributing to your ability to test the effectiveness of your project as Olivia described earlier.

And this may not necessarily include everyone who receives project activities. This could be a subset of those who receive project activities. It could also include teens in the control or comparison group as we've been talking about who may not be receiving project activities or may receive them at a different time. So there's a difference there between the evaluation sample and the intervention sample.

There's also a process evaluation sample for those who are contributing data for you to be able to describe what's happening on your project and what is being implemented. And this may also be different from your intervention sample. So it could include teens in your intervention group if you want to ask them about the project activities that they're receiving.

Your process evaluation sample could also include staff who are implementing project activities. Because you could be asking them about how they're delivering the intervention, what they're implementing. You could observe their implementation and those types of things. And you could also want to potentially include your teens or others in your control or comparison group in your process evaluation in order to learn what types of services or activities they're receiving.

Often, we learn that folks in the comparison group are actually receiving a lot of services as well even if they're not receiving your particular project or your particular curricula. So in order to interpret the results of your outcome evaluation, it's important to know what those folks are receiving as well.

We recommend working with your evaluator to look to prior evaluations as a guide in defining your evaluation sample. And by prior evaluations, we're talking about evaluations potentially of the evidence-based programs that you might be using for other evaluations that have looked at the same outcomes that you're investigating.

And we recognize that these prior evaluations may include samples and target populations that are very different from who you're targeting. And obviously, we definitely want you to know that you are the experts on your local community. But by looking to prior evaluations, you can gather information that's going to help you figure out who is going to be in your sample and how many people you need in your sample which we'll talk about in a minute.

But the types of information that you might look for include designs, evaluation designs like Olivia was just talking about; the timing of follow-up data collection, either after baseline or after project delivery; the measures that these prior studies might have used to assess outcomes that you might be looking at as well.

And then something called the effect size which is basically just how much change or how much difference between the intervention and control group the prior evaluations found on the outcomes that you're looking at. And that's going to be an important input in determining the sample size that you need which we'll get to on the next slide.

The prior evaluations can also help you make decisions about your process evaluation sample. So if you look at what kinds of data were collected and from whom in prior evaluations of a similar curricula or looking at similar outcomes might help you decide whether to include teens or staff members or other folks in your process evaluation, how to measure what your project is doing and what the activities are and what they look like and how many people are being served and those kinds of questions.

Okay. So in order to determine optimal sample size, and here we're talking about outcome evaluations, you should ask your evaluator and potentially work with a statistician if needed to conduct a power analysis. And what that's going to tell you is what kind of sample size you will need in order to detect statistically significant effects of your project if they exist.

So if your project is effective in changing the outcomes that you're interested in, the power analysis will tell you how big of a sample you need in order to be able to make that conclusion. And there are a lot of folks out there who are very experienced with doing these kinds of analyses that you can look to.

And an important consideration in a lot of projects that we've worked with in the past is being able to account for the unit of assignment or analysis as part of the power analysis. So if you do have multiple tools or multiple sites and teens or students within those groups, the power analysis will also be able to tell you how many schools or how many sites or groups you need in order to have adequate statistical power. And that's important to account for to get the right answer as far as what sample size you need.

Once you have a number from a power analysis in terms of the total sample size you need, that's going to tell you how many individuals you need to be part of your sample at the final follow-up data collection. And then it's important to consider what you expect of the rates of parental permission as well as participation, both at baseline and follow-up might be. So that you can sort of work backwards towards a number to start with in terms of who you're going to approach and how many.

So think about all the different steps where teens or individuals in your sample might drop out of the process. So you may approach a certain number of parents and some proportion of them will provide permission. And then of those, what proportion of their teens do you think are going to participate in your baseline data collection? And of those, how many might participate at follow-up? There's lots of opportunities to lose sample along the way. And so you want to be able to approach a large enough number at the beginning so that you'll end up with the right number at the end.

We always recommend erring on the conservative side for this. So if you're not exactly sure what to expect and you think there might be a range. So 70-80 percent of the teens who participate at baseline might participate at follow-up. Go ahead and assume that more conservative estimate so that you're likely to have adequate sample size.

There also may be some outcomes for which you need a larger sample than others. We know that outcomes that are particularly rare or particularly common. For instance, if you have a very young sample, sexual behavior outcomes might be pretty rare. And those actually tend to require a higher sample size in order to be statistically significant effects.

And so if you have different outcomes suggesting different sample sizes, you might also want to go with a more conservative estimate. And that, of course, assumes that you're able to do it, that you have the available communities, teens in your community or available number of schools and that kind of thing to be able to go with that larger number.

So again, we realize we're speaking to a lot of folks who are not evaluators. So if this information is not in your wheelhouse, take this information and talk to your evaluator about it and they'll be able to work with you on the power analysis.

We're learning that in some grantee projects, sampling might be appropriate. Basically, it's possible that once you look to prior studies and a power analysis, you actually learn that you need fewer respondents in your evaluation than you are planning to serve with your project.

So, for instance, the evaluation of Reducing the Risk included 758 teens at thirteen schools. It's possible that some grantees are planning to serve 2,500 teens with Reducing the Risk in schools. And so, think about how you might want to select a subset of those teens for evaluation.

Another example is the Safer Sex curriculum. The evaluation of that included 123 teens in health clinics. So that's a much smaller number than some of the school based studies. And if you are planning to serve say 300 teens with that curriculum, you may not need to include all of them in the evaluation.

Now, the reason why you might not want to include every teen who's being served in your evaluation is that data collection is very costly in many cases. It just takes a lot of resources and staff and time to collect good quality data from lots and lots and lots of individuals. And so we recommend focusing on quality of data collection, not necessarily quantity once you've reached the required number, the required sample size.

And by quality, we mean things like putting a lot of effort into staff training, saving a lot of resources for follow-up data collection which can be very costly when you're trying to track down high risk youth and others in your target population. And really honing your procedures and keeping data safe are the kinds of things that you should be saving resources for if you can.

So think about how you might want to choose a subset of the folks that you're planning to serve for evaluation if you can do that. And if you decide to do that, you can still include everybody who's being served in your process evaluation. You can collect information, for instance, about how many project activities, what type of project activities all of them are receiving. You can still include them all in your performance measures which I think is an important point. So we're just talking about in terms of outcome evaluation and lots of primary data collection which is resource intensive. Think carefully about how many you really need and whether you can conserve.

So now that we've talked through some considerations for a sample, we would love to hear from folks via the chat window once again. If you could chime in and let us know what are one or two important things for you to consider for your project in defining your evaluation sample.

And we will have time at the end of this section for more questions. Although, you're welcome to chat those in at any time. We'd really like to give you the opportunity just to share what you're thinking about at your site. Because there may be a lot of folks at cross sites who have similar ideas and things that they could share. So we're watching the chat window.

Well, based on the prior question about low numbers, the sample again. We acknowledge that some of you are working in communities where the numbers of folks you have access to is much, much lower than what your parent office might suggest or what the prior evaluations have done. And we can certainly understand that and think everybody needs to just do the best they can with what they have.

It looks like we do have a couple of questions coming in now. This says, "Will you be addressing the challenge of obtaining parental consent in order to conduct evaluation?" I think that's probably another two hour webinar. What do you think, Olivia?

DR. OLIVIA ASHLEY: Absolutely.

MEGAN: So Olivia will be talking about IRB considerations which may include parent consent. We won't be going into a lot of detail about that on this webinar. But we are certainly available for technical assistance with parent consent and there will probably be other presentations and information provided about that and other opportunities.

DR. OLIVIA ASHLEY: I just wanted to add that that is a major consideration for defining your evaluation sample, if you're working with an IRB that requires active parent consent, difficult to put together an evaluation sample if you cannot convince parents to let their students participate in your evaluation.

So I think that that's actually a really strong observation about something that may define your evaluation sample for you. And we're going to talk a little bit about TA at the end. If you would like to talk more about all the interesting things that we've tried with parent consent and the successes and lessons learned that we've had, we're happy to share with folks individually if you'd like to make a TA request or your project officer.

MEGAN: Another question that has come up is, "What are you typically using as your population measure? So that can be all students in the state, all students in the project, et cetera." That's partially going to depend on how you decide to recruit and collect your sample for your evaluation. And often with projects that we've worked with in local communities, the population really is sort of defined by either the community or the project or a smaller parameter than say the state.

As Olivia was mentioning, we know that you all are working with target populations and target groups that don't necessarily look a whole lot like the larger state or even necessarily your county or other things like that. And so we don't necessarily expect that your results will be able to be generalized to the entire state or a large group like that or even the nation.

So think about when you're deciding on how to define your sample what those folks are going to look like in comparison to either your larger community or possibly if it's just going to be limited to the teens or students that you're serving on your project. But we rarely see projects that are able to really generalize to a very large population like the whole state or all youth that have particular demographic characteristics. Usually, the conclusions are just made about the teens that were served on the project or perhaps teens in the schools or communities that you're working with if your sample looks like the larger group of teens in those schools or communities. Do you have any to add to that, Olivia?

DR. OLIVIA ASHLEY: Sure. I'm thinking about an evaluation that we did where we had about 2,200 students in I think 24 schools. Half of those were comparison groups. Half of those were treatment schools. And then there were many other schools that were getting the program. And there were also many other schools that were getting the program. And there were also many other students within the schools who were getting the program who were not in our evaluation.

And so one of the things that we did was that we looked at the demographic characteristics which was all that we could learn at that time of the schools where we were getting our evaluation kits from. And we were able to work with the statisticians to weight the data so that we could look at how similar or different our evaluation participants were from the larger school population. And the statistician was able to weight the data so that they could look at what the responses would look like if we had had the same distribution as the students in school.

So if there were some underrepresented students whose parents would not provide consent, they may come from families who don't have a lot of parent involvement or parents who are working several different jobs or who are not comfortable with federally funded evaluations or research for whatever reason, we could consider what our data would have looked like if we'd had the same distribution in our evaluation as we did in the school. So that's one possibility if you have some data about what your larger population looks like and you think that it's something that is of interest to you to try to measure, you could ask your statistician or evaluator to weight the data for you.

MEGAN: More great questions. We're going to go ahead and keep moving. So we've talked over the last two sections of the presentation about design and sample issues. And FYSB was interested in us sharing some tips for working and evaluating your local evaluator given that a lot of the folks on this call are not evaluators, but are program administrators or program staff.

So a lot of these things are very similar to what Olivia identified as tips for finding a good local evaluator. On the flip side, if you're working with a local evaluator and they have never successfully recruited a comparison group or do not know how to identify a comparison group or how to advise your staff about recruiting one and do you want to implement a design that involves the comparison group. Or if they don't know what a power is for determining needed sample size for your evaluation. Those might be suggested things to look for to indicate that your evaluator might be in over their head or just might not be able to provide the support and the evaluation services that you're looking for for your project.

In some cases, technical assistance can help with that. And in some cases, that might not even be enough to help local evaluators with these pitfalls. So if you think you're in this situation, it's something to talk with your project officer about. And we want to make sure that you have your evaluation needs met and that's FYSB's interest as well. So that you can do the most rigorous local evaluation possible.

So just to summarize the information about defining your evaluation sample, once again not everyone who receives project activities necessarily needs to be in the evaluation and the evaluation might even include individuals outside of the folks in your intervention sample.

And definitely work with your evaluator and a statistician as you conduct a power analysis to determine optimal sample size at final follow-up. And that way you'll have a sense of whether the folks you have access to are going to be enough. Or whether you need to think about other groups or communities you could reach out to or how to subset down if you're working with a very large group and serving a large group. And you don't need all of them in your evaluation. And that will help you make a plan.

Are there any additional questions or comments about things that we've talked about so far before we move on? Go ahead and chat from the chat window. Okay. Well, I'm not seeing any more questions coming through. But as we're talking, continue to send questions and we'll pause to address them as we go.

So the next topic that we wanted to share some information about is ensuring appropriate measures. And we could also probably spend a whole webinar about measures. But we have some TIPs and some considerations for choosing and developing appropriate measures. So as we were talking about with samples, it's recommended that you look to prior evaluations of the evidence-based curriculum that you might be using to find out what measures that project had impact on, where they found significant impact. So if you're using a program, for instance, that was shown to be evidence-based because it reduced initiation of sexual behavior or it impacted condom use frequency, things like that, those are the measures that you also want to include in your evaluation in addition to plenty of others. But be sure to at least include the measures for which th8ere were significant impact in prior studies.

The other place to start in choosing appropriate measures is your logic model and your project objective. So the figure on the slide is a very, very simple, probably one small sliver of a logic model that might have some things in it, some outcomes that you're looking at too.

And the purpose of this is just to illustrate that if these were a part of your logic model, you would want to make sure you have measures that address all of them. So not only the outputs that you would be measuring through your process evaluation of what is being delivered. So in this case, how many classes are being delivered. But also make sure that your measures address your short-term, intermediate and long-term outcomes. So increasing knowledge, decreasing initiation of sexual activity and decreasing pregnancy rates. You could measure all of these and you would want to in order to address the objectives that you've set out for your project and be able to actually test your logic model and see if the classes and the curricula and the activities that you're delivering are actually impacting the different outcomes.

In many cases, the measures that you might be interested in using probably already exist and have been used in the past. I know, for instance, several folks maybe interested in looking to items in the youth risk behavior surveillance or YRBS data. And it's very common to adopt existing measures for these types of evaluations.

We definitely want to highlight considering cultural or language issues when you're adopting existing measures. We have a lot of folks on the call that are working with very unique populations where these measures just really haven't been used before. And so it's important to think about how the folks in your population and in your sample might interpret different questions and the language that's going to be required to get these concepts across.

Similarly, you'd definitely want measures to be developmentally appropriate. And so, looks at where these measures were developed and who they were used for. So if they were used with high school students and you're serving middle school students, think about some terminology or behaviors or things that you might want to change in the measures to make sure that they're appropriate for the age of folks that you're working with and that you're including in your evaluation.

One example that we often worked with is the term contraception. It's just very complicated and not something that younger teens might be comfortable with. So we've reworded that term and not used it in surveys of young teens. And we talk about "ways to keep from getting pregnant" instead. So think about that.

And similarly, think about the reading level of the sample that you're working with for your evaluation that you're collecting data from and make sure that the reading level is appropriate as well. So this may mean breaking items down into really simple words and lots of statements instead of one really long run-on sentence and things like that in terms of if you have a survey that folks are going to have to read.

The other thing in adapting measures is to make sure that the schools or administrators or leaders in your community are comfortable with the data that you're collecting in terms of what you're asking, especially of teens. We found that in some communities, schools or administrators are not comfortable with asking about certain behaviors on surveys. And so make sure that the surveys and data collection methods that you're developing are accessible to them. And also, it's an opportunity to find out what they're interested in learning about your sample. And you can offer to include that as part of your measures.

In terms of instrument development, some tips that we found useful for folks, include the use of skip patterns and grouping like items together. So, a skip pattern would just be where depending on the answer to a question, the respondents are guided to skip the questions that follow. And this often comes up for us doing teen pregnancy prevention evaluation in that lots of respondents in your sample may not have dated and may not have had had sex.

And so questions about condom use, for instance, may not be relevant for those folks. And so instead of letting everybody in your sample try to answer those questions when it may not be relevant for them, building skip patterns into your survey is a great way to reduce the burden on them and also not try to lead them down the path of thinking that they should be using condoms for some reason if they've never had sex.

Another thing that we've considered in the past is trying not to put your most important outcomes at the end of a survey in case there's a chance that some of the teens won't finish. There may be exceptionally slow readers or just kids that get distracted or fatigue out. And if there's something that you really want to make sure that you have data from everybody on, it's best not to put that at the end. Although, it's also best not to start out with very sensitive questions. So somewhere in the middle is usually what we try to target for the really key outcome measures.

In any instrument development effort, it's really important to gather input from your target population, especially if you're working with a unique population where these measures haven't been used before and you're not really sure what considerations to think about in developing appropriate measures.

Doing formative research with your target population, even if it's just a handful of teens that you get to look at some items and give you some input. And then pilot testing is a great way to assess whether your respondents are going to be able to understand what they're looking at, whether the instrument will yield consistent results across time or across people which is what we mean by reliability. Whether the questions and data collection is measuring the things that it's supposed to measure which is what we mean by validity. And then just how much you're burdening folks.

So how long it's going to take them to complete the data collection and complete the instrument and how tiring it is. And what we've often done at the end of pilot test surveys is ask some specific question about whether it took too long or was just the right amount of time, whether the questions were hard or easy to understand and ask for specific feedback about that. And whether the teens were being honest in their answers is another great pilot test question to find out whether you're likely getting accurate data.

Another consideration in choosing and using appropriate measures is thinking about what the timeframe reference is when you're doing to follow-up data collection. So what are folks thinking about when they're answering questions in terms of time? And so one tip is not to ask respondents to try to remember too far back about their own behaviors. Because you're likely to get more accurate data if they're not having to remember too much more than a month or a few months back in time.

But really the other thing that we want to talk about in terms of timeframe reference is about what timeframe is included in the follow-up measure in terms of when your baseline data collection was and when your intervention is delivered. So if your follow-up measure includes the period during your intervention or especially if it includes the period before your intervention began maybe hard to capture change from baseline to follow-up and attribute that change to project activities.

So an example of this is included below. So if you use a question of "Have you had sex during the past three months," well, if you're asking and it's only been a month since you did a baseline data collection and started your intervention, a respondent could answer yes. But maybe they actually had sex before the baseline data collection and before your intervention began and you're then not really capturing any effects of your intervention with that measure.

So that's what we mean when we talk about the period of time when your measures are referencing. So just be careful about that and how you're wording that depending on the length of follow-up that you're using.

Another tip to be aware of is that if you want to try and be really specific and really zero in on the time period since the end of your project, so since program delivery ended and try to get information about behaviors or attitudes or things in that window of time, that works great except not with respondents in your comparison group or your control group who haven't actually received your project activities because they probably won't know what you're referring to with that. So just another tip.

The other example at the bottom of the slide is about annual school data on dropout rates if that's something that you're interested in collecting, just a heads up to think about how that would align, which years of data, for instance, would align with your project activities and whether you need to actually collect multiple annual years of data in order to capture the change that may be occurring as a result of your project implementation.

So just think about along the lines of your design and how you're coordinating project implementation with the evaluation, think about that in terms of measures as well of what you're asking people to think back to when they're reporting on behaviors or other outcome measures. And keep your expectations about how much change you're going to see realistic. We know that effects on sexual activity take time to occur and to measure.

And this is even more the case for things like pregnancy and birth rates. It just takes time for these outcomes to show up, especially if you're working with young populations where the rates of sexual activity and pregnancy are low. It's going to take a while before you can actually see change and see differences, especially between your intervention and a comparison or control group on those.

So there are some options for looking at shorter term outcomes, intermediate outcomes, especially in your early follow-up data collection. So if you're doing a follow-up data collection, just a few months after baseline or after your program begins, there are some alternatives to what might be the more desirable outcomes. So, for sexual intercourse if that's one of your outcomes, some possible shorter term outcomes would include attitudes, intentions and pre-coital behaviors that don't include intercourse.

If you're looking to impact pregnancy rates, some shorter term outcomes for the early follow-ups should include motivational and self-efficacy to avoid pregnancy, use of contraception, intentions to use contraception. And these can all be drawn from your logic models and your project objectives in terms of what the shorter term and intermediate outcomes are that you're trying to impact on the way to getting to sexual behaviors and pregnancy and birth rates and those types of ultimate outcomes for your project.

The other thing to think about in terms of realistic expectations about change is that a lot of normative change in teens is in the direction that is the opposite of what you're hoping to do. So, for instance, teens are more likely to initiate sex as they get older. And so, it's probably impossible that your project will yield, you know, fewer teens have ever had sex that follow-up the baseline because that's just not possible. Once they start, they're going to have ever had sex.

So think about what's realistic as far as perhaps a lower increase in initiation among your intervention group and your comparison group. Teens' attitudes tend to get more accepting over time and their intentions tend to increase over time in terms of these risk behaviors. And so it might be more realistic to think about smaller increases in those things in your intervention group versus the comparison group, rather than decreases necessarily. The strongest program often don't reverse developmental trends completely. So something to be aware of.

We talked a lot about looking to previous evaluations of program that you're using. And another way to look at this is to guide the timing of when you complete follow-up data collections. So, for instance, if the original evaluation of a curriculum that you're using showed effects on unprotected sex at six months, but those effects disappeared at a twelve month follow-up, be sure to measure that outcome at six months and don't necessarily expect effects on it at twelve months.

And also think about other outcomes that you might want to measure. Because if the program's effects on a particular outcome disappeared over time, then you might be better off hanging your hat on a different outcome measure. But it's still important to measure the things that were impacted in prior studies.

Also, work with your evaluator to consider differences in age in terms of the sample that was in the original evaluation versus your target population. Because if you're working with a younger population, then the sample that was evaluated in a prior study, the prevalence of certain behaviors may be different. And so you may need to test outcomes over a longer period of time, as I was mentioning, to sort of let those behaviors unfold over time and increase in prevalence in order to be able to see an impacts of your program.

We're getting lots of questions. But first, we have another quiz and we'll be ready for questions in just a minute. So the scenario here is that your evaluation will test the effects of an evidence-based curriculum among middle school youth. A previous evaluation of the curriculum with high school youth showed effects on sexual initiation. How can you ensure your measures are appropriate? I'm going to read the options and then we'll go ahead and open the quiz.

So option A is you can survey youth about whether they've ever had sex immediately before curriculum and implementation and a one-month follow-up. B is survey use about whether they've ever had sex immediately before curriculum implementation and at six month follow-up. C is survey use about intentions to have sex immediately before curriculum implementation and at six months follow-up. D is to ask teens during a focus group whether they have had sex. And E is both B and C. That's not complicated enough for you.

Okay. So we've opened the poll and again we only have one minute. So go ahead and register your responses and then we'll talk about the answers. Okay. The quiz is closed. We're just waiting for the responses to load up and then we can share them. It looks like a lot of people may have still been pondering the answers. So we'll talk about them in just a second here.

So it looks like of the folks who responded, many, many, many said E which is both B and C and that is in fact the correct answer. So the most important thing to do in terms of ensuring measures are appropriate is to use a similar outcome as what the previous evaluation showed which is whether youth have ever had sex and a six month follow-up which is in B is better than a one month follow-up which is an A, just in terms of being able to see effects on sexual initiation unfold over time, especially among this young sample. But the reason that C is also a good answer is that because you have a young sample in this situation, it might also be important to measure intentions in case you have such low rates of sexual initiation that you're not able to see effects.

So we wouldn't recommend only surveying them about intentions, but maybe doing B and C together. It looks like a lot of people were on that page. And the next most popular answer it looks like was intention, to survey them about intentions, which I'm guessing is because people are thinking about the young age of these teens. And certainly, if you have constraints against measuring certain sexual behaviors, intentions is a fallback. But again, in terms of ensuring appropriateness of measures and giving yourselves a good likelihood of having some exciting outcomes, we recommend at least including the measures that were in previous evaluations of the curriculum. So a great job on that quiz.

So just to summarize the information about ensuring appropriate measures, as I was just mentioning, consider measures that your selected curriculum has been shown to improve. Make sure to attend to reading level and other developmental issues in your population as well as just other characteristics of your target population to make sure that they're really understanding the measures and the data collection approaches that you're using.

Make sure not to send an inadvertent message that everyone has had sex or everyone has dated. And so be sensitive to teens who may not have engaged in these behaviors by building in skip patterns and developing your measures. So that they don't have to answer a bunch of questions about things that they've never done.

And perhaps the most important thing is to do a pilot test of their instruments that you're developing so that you have the opportunity to improve them and make them really usable and relevant and yield high-quality data before you start your main evaluation.

So I believe a lot of questions came into chat while I was talking. So we'll take a look at those now. We have one that says, "We are working with a small, well-connected population. What can be used to reduce or eliminate contamination between the project group and the comparison group?" This is going back to our discussion of evaluation design. You want to take this one, Olivia?

DR. OLIVIA ASHLEY: Sure. I had the benefit of reading these questions and thinking about them while Marnie was having to present. So I did think about some answers to those. I came up with three options and there may be other options. One is we talked about a potential comparison community. And I know the comparison community may not be exactly like the project community. But if your community is that small and that well-connected, it may be very difficult to avoid contamination. And you may want to think about another small well-connected community to match to serve as your comparison.

Another thing that we have done when we've just had to deal with contamination. I mean, we've been on studies where we've had schools right across the street from each other. And for one reason or another, they needed to be a treatment and comparison school. And these kids are congregating or having lunch with each other or that kind of thing. We at least setup some sample inclusion and exclusion criteria for the worst kind of contamination.

So, for example, we made sure that no students were in the same household if they were in our evaluation and certainly not a brother and a sister or two sisters, one in the treatment group, one in the comparison group. So you have as much close quarters and talking and sharing of values and that type of thing that you could possibly have going on.

And then another thing that we've had to do when we've had to have contamination, we knew it was going on. We had to proceed with it. There wasn't much that we could do about it. And that was that we measured it. And so what we did is we made sure that the project group and the comparison group were defined based on something that was easily described.

And so it could be that you've got two schools in your community, let's say two middle schools, School A and School B. And so you could say "at follow-up" on your questionnaire, among the students who are in your control group, they may be your School B. You could ask them have you talked with other teens at School A about sex or contraception which could be "how to prevent pregnancy" if they don't know what contraception means.

You could ask them have you talked with them about the name of your program like Making Proud Choices? And we have found that when we've asked that question, much like when you ask people, "Have you ever seen a UFO?" you get a lot of people who say yes. But they didn't actually do that. So there's a validity problem.

We've also thrown in a decoy question. So we'll ask, "Have you talked with them about Making Proud Choices? Have you talked with them about considering Healthy Decisions in capital letters?" And you can ask your statisticians to do some controlling. It's almost like the live scale on the MMPI for folks who are familiar with that. If they were liable to answer falsely about a program that doesn't exist, then you might want to discount their answers to the program that does exist.

If you only have one school in your community and you really are concerned, you could try to divide your project and comparison group by one group receive second period health class and one receive fourth period health class. And then you could ask your comparison group have you talked with teens who are in fourth period health class or who got fourth period health class last semester or last year, that type of thing.

So it needs to be something that teens can easily report back to you about some kind of communication they may have had. You can ask them, “Have you ever dated anybody who was at School A?” That might have some bearing. So, those are my thoughts on it. I don't know if you’ve got other thoughts on that, Marnie.

MARNIE: No, great strategy. We received another question that says, “Are effect sizes required for para-analysis? If not, what do you use?” In my experience, we’ve always had some kind of estimate of how big of an effect we’re expecting because that’s sort of a key driver of what kind of sample you need. The larger your program effect, like if you’re going to make a huge reduction in sexual behavior, you need a lot fewer teens in your sample.

If you’re expecting a very, very small effect, which often behavioral interventions are expected to have fairly small effects, then you need a lot larger sample so that the differences between them won’t be as large. And so you need a lot, a lot of teens for that. I haven’t ever seen a para-analysis that didn’t include some kind of estimation of how big of an effect to assume. And effect size estimates are not always very accessible from prior studies.

Sometimes you have to look around a little bit and there are different types of analyses that are used for percentage variables, variables that are dichotomous and give you percentages. So the percent difference in pregnancy, for instance, versus continuous scales. Those are going to give you different information. But those can both be applied to a para-analysis.

So definitely ask a statistician. We’ve often sent articles, literature, to our statisticians to have them figure out what information needs to be extracted from it. So you might be able to do that and try to get some input about what needs to be pulled in. Anything else to add on that?

DR. OLIVIA ASHLEY: I guess I agree with everything that Marnie said. Just two things for those of you who are statisticians, I’ve not had to actually use a [inaud.] effect size. As Marnie was saying, it could just be a study that showed what the baseline percentage was for the comparison group and the follow-up percentage for the comparison group, the baseline percentage for the treatment group and the follow-up percentage for the treatment group. And the statistician's been able to figure out what needs to go in the para-analysis from that.

The other thing is that we can offer consultation by our statisticians if you want to request TA. We can't do your analyses for you. But we absolutely, if you're trying to do a power analysis and you have an evaluator, a statistician, who has questions like that, we're happy to draw in a statistician who would know way more than Marnie and I would about options to talk with one of your staff if you want to make a request to your project officer who may think that that's appropriate.

MARNIE: Thank you. And often, when we don't have an actual size estimate when it's percentages or something like that, we're often asked to give input about how much of a range we think there might be in the sample on those measures and they usually need some kind of measure of variability that's expected to be able to hone in on a sample size calculation. It looks like we have some repeat questions. And then another question that came in is, "What do we do when schools and community organizations are not comfortable with or resistant to asking youth about the required performance measures?"

Well, there are options for potentially existing data which you may or may not have in your community or school. Do you have any other thoughts about that, Olivia?

DR. OLIVIA ASHLEY: Sure. I think this person is referring to required performance measures by FYSB for the multi-component evaluation. And so I guess I'll throw out some thoughts on that and then FYSB may have something they want to add to this.

One thing that we've seen when we've worked with lots of other grantees, and this may not be what this person's experiencing, but we have had people who've anticipated resistance. But when they actually asked, there wasn't resistance. And we've had that with IRBs. "I'm sure my IRB would never go for that," or, "I'm sure the school will refuse." So we recommend that you not make an assumption if you haven't already ask, go ahead and ask. You might be surprised at the answer that you get.

And then we've seen some guidance that if the concern is about age of students that performance measures for sensitive questions like about sexual behavior may not be required for students who are younger than age fourteen. So that may be something to talk with your project officer about. And I also want to point out they may not be comfortable with some of the sensitive performance measures questions. But that's not all the performance measures questions. There may be many others that they are comfortable with. So not to throw the baby out with the bath water.

And then as you're talking with the project officer, my experience is that FYSB has been very sensitive to community norms and politically sensitive issues and ecological considerations. They may be considering whether they can offer a waiver for that situation. They may not be able to offer a waiver. And to my knowledge, there's no decision on that yet. But that may be worth a discussion with your project officer just to make sure that the things that you're trying to do are what FYSB's expecting. And so I'm wondering if Sarah or LeBretia or someone else wants to add anything to that.

LEBRETIA: Hi, everyone. This is LeBretia. Not particularly at this point. I think your answer is sufficient, your response to that.

DR. OLIVIA ASHLEY: Great.

MEGAN: Do you want to take one more question or should we wrap?

DR. OLIVIA ASHLEY: One more and then we need to move on so we can wrap up.

MEGAN: We did receive one more question that says, "Is there any concern about how to use the skip patterns we identify among the students taking the survey which students have or have not had sex? For instance, if you ask those students who have had sex, fifty questions and those who have not had sex, ten questions as an example, you may be inadvertently outing those students who are sexually active." I think that's a great point and something to be very careful about in designing your measures.

So you want to think about how you can set up the surveys so that forty of your fifty questions aren't about things that are related to sexual behavior that you have to have had sex in order to answer. And I can imagine that there are probably a lot of other items that you can ask of everybody that aren't necessarily about the skip pattern or after the skip pattern.

So think about what you want to learn about everyone and then see if you can find ways to build in the skip pattern possible in a place within the survey. So that kids who are skipping over it still have questions to answer after that. But I think the bottom line is you don't want that much of an imbalance between the students who are skipping out of questions because they've never had sex and the kids who need to answer all the questions because they have. And it's a really good point to make and share with everybody. Do you have anything to add, Olivia?

DR. OLIVIA ASHLEY: No, I agree.

MEGAN: Okay, great. Well, I think we're going to move on. It looks like there might be another question and we can try to find time either at the end or take this as a note for later and try to get back to folks about additional questions. We just want to make sure we get through at least this last little part here before we wrap up today. So I'm going to turn it back over to Olivia for the remainder of the presentation.

DR. OLIVIA ASHLEY: So this last part is about working with an institutional review board or IRB. And there are many folks on the call I know who have a lot of expertise on IRBs. They may actually be a member of the IRB. So I'm sure there are folks who have lots of detail that they may be able to add here and serve as a resource. So I just wanted to make sure that everyone is familiar with what the purpose and the role of an IRB is and that is to ensure protection of human subjects who may be participating in an evaluation or another type of research or study.

And the IRB reviews protocols and materials for evaluations that involve human subjects. They do report to the Office for Human Research Protection which is part of the federal government and that is within DHHS. So there are DHHS regulations that they have to comply with. And their job is to consider what is ethical in human subjects research.

And certainly within our country at least, there have been large precedent of unethical human subjects research that has been conducted and that is why it's very important for IRBs who are reviewing materials to consider that and prevent that from happening again in the future.

This graphic here is from the Belmont report you can see at the bottom, which was a report in the late ‘70s about ethics. It’s a really important report about protecting human subjects in research or evaluations. And there are three main principles in this report.

The first one is beneficence which means that the research or evaluation should not harm anyone. Teens or schools or parents or whoever is participating in the evaluation should not be worse off because they were in your evaluation than they were if they had not been in your evaluation.

The one on the right is respect for persons and that is that people need to be able to make voluntary decisions about whether they want to participate in their evaluation. No one should be forced or coerced or threatened with some outcome that they don’t want or offer something that is so big that they could not actually refuse that, particularly if they’re working with an impoverished population or disenfranchised population.

And then the last one at the bottom is social justice. And that is that everyone in your target population should have the opportunity to participate in your evaluation. So it should not exclude certain people in your target population. They we’re only interested in collecting data from some people in our target population. Or that we are only collecting data from some specific population and not opening up the evaluation to all types of people that we are serving in our target population.

This slide is about responsibility. And ultimately, grantees are responsible for making sure that evaluations that involve human subjects are conducted in accordance with fairly broad federal regulations that just make sure that human subjects are not harmed as part of evaluation research.

But grantee staff and subawardee staff do share responsibility for protecting the rights and welfare of human subjects and for complying with any federal or institutional regulations that may apply. And institutional regulations maybe things that you as a state or as a member of a university or a school system or a public agency, you may have your own regulations that you’re familiar with that you need to comply with.

IRB review may be required for local evaluations to ensure that human subjects are protected. And just a heads up, you may be aware or may not be aware that subawardees, for instance, if you're working with schools or hospitals or other organizations, they may have their own review process that may be separate from any decisions that you're making about IRB.

So there are at least three possible levels of IRB review. There may be many more than these in some IRBs, but these are the three main ones that we've seen. And that is an IRB may decide that a certain type of data collection is exempt from IRB review because there's minimal risk or the data collection is anonymous.

And minimal risk, we tried to put a definition here. What does that mean? It means that the chance of harm or discomfort or the amount of harm or discomfort is not anything greater than what someone would encounter in their daily life or a routine physical test or routine psychological examination.

An expedited review also could be something that's required and that means that the IRB is willing to consider your application on a fast track. Or they may not require quite as much paperwork as they would if you do not have as much risk that's involved for the people in your evaluation. And a full committee review is when they want to see everything. It's sort of the most scrutinizing that an IRB could possible be.

And so typically, the level of review is determined by the IRB. We've just not been in any situations where we've been able to tell our IRB, "We think this is exempt and therefore it will be exempt." Our IRB tells us. And we frequently pick up the phone and ask that question in advance. But we don't get to make that decision typically.

And so we think the first step is to work with your evaluator to identify an IRB that you may want to consult. There are IRBs at universities. There are IRBs that are independent where you can pay a fee for someone to oversee your evaluation. And certainly, there are research institutes that have IRBs.

So we recommend that you ask your evaluator to contact an IRB and just have a quick conversation. "What are your evaluation procedures? Are you surveying students using the Internet? Or are you going to collect paper and pencil? Or are you going to call them? Or are you going to do a focus group? Are you going to look at attendance records as part of your implementation evaluation? How old are the participants in your evaluation? And who are they? Are these students who are in the juvenile justice system or in foster care? Or are these students in a normal school population?"

And then ask about what the level of review is that might be needed and what materials are going to be needed to prepare, what their requirements are going to be about consent or assent as well as how long do they think a review is going to take.

And one of the things that we've done is that we've not waited to get started after we've submitted every possible thing that is going to happen on the study. We've frequently submitted, "Well, we just want permission to do our pilot test right now. Or we just want permission to collect our process evaluation data. And we'll come back to you when we're ready to start looking at our data collection from students. Or we were just looking today at what we're going to do in terms of collecting data from parents."

Typically, when you're looking at an IRB submission, they want a full description of the study procedures, your consent forms or scripts, any recruitment materials like flyers or lead letters or ads you might be posting. And what do the data collection instruments look like? And the IRB will tell you what they're looking for. And there is an OHRP website that may have some useful information for your evaluator to check out.

This is a list of the kinds of things that the IRB may be looking for, what level of risk there is. How are you going to make sure that answers are private? Are you planning on getting consent or permission or assent? And then what does the consent form contain in terms of what you're going to disclose to folks so they can make an informed decision?

I'm not sure that we're going to have time in the remaining two minutes for a quiz. So I think we're going to skip this one if that's okay. I'll give you a hint about this though if we'll just back up. And that is that part of this is what is always true. And so, our thinking in developing this quiz is that every IRB has their own requirements. And so don't make assumptions about what is always true. Those of you who stood on the IRB, you know what's always true for your IRB.

But every IRB is different. So consider that you may want to ask questions about what's needed rather than making assumptions about that.

All right. We've got one minute left. And I'm seeing that there are several questions and comments. Marnie, is there one that you're seeing that you think we should try to address before we wrap up?

MARNIE: Well, there are a couple of comments that came in about measures. But there is one about IRB review in tribal communities. So it says, "For Tribal communities, the Tribal IRBs also consider potential damage to communities since there are many past and present. The IRB reviews would be in addition to or instead of the usual IRB review. And depending on a specific IRB, pregnancy and sexual behavior questions may be seen as too sensitive, please be prepared to work on a data sharing agreement to protect individual community data from being released without the permission of the community involved."

DR. OLIVIA ASHLEY: It's an excellent question. It maybe something that FYSB says we want to think about. And just in case everyone doesn't know, things like performance measure data are going to be aggregate. But this is a really strong consideration, that even if it's aggregate, we're not talking about individual data. We're talking about how a community is represented. So I think this is an important thing to consider, for FYSB to think about and for Tribal communities as well as [inaud.] communities that folks are working with.

One of the things as evaluators when we work with project implementers is we think a lot about "do no harm" and publishing results that are linked to the identity of a school, a community, a district, a neighborhood. Even if we think it's a huge improvement, that information may be considered private and sensitive to that community or that school. And so we are very sensitive about checking in and making sure that we are not disclosing rates of sexual activity among teens within a particular group that's identified because we don't want to harm anyone's reputation or the way that they are perceived in the larger community. It's a great point.

All right. So why don't we cover some take-home points and then let folks go?

FS:Can I just interject? I know that we have to end, Olivia, because we're a little over time.

DR. OLIVIA ASHLEY: No, go ahead.

LEBRETIA: I would say that for those questions and comments that were not addressed today, we can definitely take those under advisement and provide feedback to grantees on some responses to those, including the key question that was raised about the native populations. And we can share that through project officers directly with grantees or through our next quarterly grantee's meeting.

DR. OLIVIA ASHLEY: Excellent. Thank you, LeBretia. That's a good point that we'll hang onto these. I see really good points, comments and questions that I think everybody could benefit from more discussion about. So our take home point today, work with your local evaluator. Plan for the most rigorous outcome evaluation design that's possible for you.

Define your evaluation sample to maximize your resources and results. Keeping in line with the comment that we saw for folks who are very concerned about saving as much money as possible for project implementation.

Don't survey every single person that's receiving your project if you don't need to do that. You want to be good stewards of federal funds and look at what is required for your power analysis and half of those in your power analysis are going to be comparison students. So you definitely do not need to track every single person receiving your project if you have a large project sample.

Invest in high quality developmentally appropriate measures that help you answer your specific evaluation questions. We recommend that you talk to an IRB early in the process about what's needed and to help you plan for review and reach out to your project officer if you want technical assistance about any of the topics that we've talked about today, any of the comments or questions in the chat box or any other issues that are related to project evaluations.

And our last slide is some additional resources for folks who may want to learn more about evaluation.

We really want to thank folks for their very thoughtful comments and questions today and for staying just a few minutes over. And we look forward to seeing what great evaluations folks are embarking on. Thank you very much.

(END OF TRANSCRIPT)