Description of Audio Data Collection for Program Evaluation

Good afternoon everyone and thank you for joining us. The presentation today addresses data collection for program evaluation. This webinar is about one hour long and I will stop a few times to take clarifying questions and hear a bit more about some of your experiences with data collection. In addition, we will have about 30 minutes at the end of the webinar to get into more in-depth questions you may have.

My name is Andrea Robles and I am a Research Analyst with the Office or Research and Evaluation at CNCS.

For this presentation, we have identified a number of learning objectives.

By the end of this presentation, you will be able to:

- Understand key questions to consider prior to selecting a data collection method
- Understand the importance of selecting appropriate data collection methods
- Describe some of the advantages and disadvantages of data collection methods
- Understand the difference between quantitative and qualitative methods and their roles in process and outcome evaluations

I just want to point out that this webinar is more of a conceptual overview of data collection and it is not intended for you to become experts in how to collect data – so no worries there - but it will certainly give you an understanding of the components necessary to think about data collection methods and how they may be used in different types of evaluations. Also, we know all of you are at different levels with your knowledge of evaluation including data collection, so I want to state in advance that we look forward to hearing from all of you and all questions and comments are welcome.

We begin this presentation with key questions to consider prior to selecting an appropriate data collection method.

There are four key questions to consider prior to selecting a data collection method for your evaluation. These include:

1. What is the purpose/objective of the evaluation?
2. What are the research questions?
3. What is the type of evaluation design?
4. What resources are available for the evaluation?

We will discuss each of these questions in the following slides.

Each evaluation should have a *primary* purpose around which it can be designed and planned, although it may have several other purposes. The stated purpose of the evaluation drives the expectations and sets the boundaries for what the evaluation can and cannot deliver.

In defining the purpose of the study, it is helpful to identify why the evaluation is being done, what you want to learn from the evaluation findings, and how the information collected and reported by the study will actually be used and by whom. Examples include producing evidence that a program is meeting its intended outcomes or understanding how to operate the program more efficiently or identify barriers to implementation.

In general, defining a specific purpose for your evaluation will allow you to set parameters around the design you use, the data you collect and the methods you will use.

Questions about why your evaluation is being done and how the information will be used should be discussed among a variety of program staff, and any other individuals who may be involved in the evaluation to ensure there is consensus as to what the evaluation will accomplish.

As you begin to define the purpose and scope of your evaluation, it is helpful to refer to your program's logic model to clarify and confirm your program's operations or processes and intended outcomes. For an overview of logic models, CNCS grantees can refer to the module, "How to Develop a Program Logic Model" located on the Knowledge Network.

To briefly review, a program logic model is a detailed visual representation of your program and its theory of change that communicates how your program works, the resources you have to operate your program, the activities you carry out, and the outcomes you hope to achieve. It communicates how your program works by depicting the intended relationships among program components. Key program components consist of:

- Inputs or resources - which are considered essential for a program's activities to occur.
- Activities – which are the specific actions that make up your program or intervention.
- Outputs – what a program's specific activities will create or produce, providing evidence of service delivery (e.g., the number of beneficiaries served or the number of children improving reading scores).
- Outcomes - the specific changes that may result from a program's activities or intervention.

In addition, we can think of a logic model as essentially having two "sides." The **process** side focuses on a program's implementation or its planned work – inputs/resources, activities, and outputs (direct products). The **outcomes** side of the logic model describes the expected sequence of changes that the program is to accomplish, which can be short-term, medium-term, and/or long-term changes. The outcomes side reflects the difference the program intends to make.

Please see your handout, for an example of what a logic model might look like for a fictional AmeriCorps homelessness prevention program.

The next question to consider prior to selecting a data collection method is what are your research questions? Your evaluation's focus on program processes or program outcomes will generate different kinds of research questions, which in turn, will determine your data collection methods. As this graphic illustrates, process evaluations address questions about program operations, namely the who, what,

when, where, why, and how many of program activities and program outputs. On the other hand, outcome evaluations measure a program's outcomes and assess program effectiveness.

These research questions, in turn, determine the data collection method that will be used. And I will describe this some more in the last section of the webinar.

Once you've established the purpose of the evaluation and identified your research questions, the next step is to decide which type of evaluation design you will employ. Just as there are two sides to a logic model – a process side and an outcome side – there are two common types of evaluation designs: a process evaluation design and an outcome evaluation design. The side of the logic model that the purpose and scope of your evaluation focuses on, will largely determine which type of evaluation design should be used for the evaluation.

In this table we highlight some notable differences between process and outcome evaluations, especially in terms of their goals. For a process evaluation, findings are most often used to change or improve the program. A process evaluation can be used to document what a program is doing and to what extent and how consistently the program demonstrates fidelity to the program's logic model. An outcome or impact evaluation can be used to determine the results or effects of a program. These types of evaluations generally measure changes in program beneficiaries' knowledge, attitudes, behaviors, and/or conditions thought to result from the program.

To answer the types of research questions associated with a process evaluation, a comparison group is generally not necessary. While not required, the more rigorous outcome evaluations include a comparison group against which to measure changes in program beneficiaries. These types of outcome evaluation designs are referred to as impact evaluations. The use of a comparison group provides additional evidence that observed changes in program beneficiaries were due to the program or intervention. Thus, impact evaluations are better able to measure or estimate the **impact** of the program on beneficiaries.

Just as research questions determine the data collection methods that will be used, so do the type of evaluation design. And again, I'll show examples about this later in this webinar.

It is important to remember that CNCS has different evaluation requirements for large and small recompeting grantees in terms of which evaluation design they may use to assess their programs. Large grantees are those receiving annual CNCS funds of $500,000 or more. Small grantees are those receiving annual CNCS funds of less than $500,000. You should note which type of design is required for your program. The required design will drive the types of data collection methods available to you. For more information on evaluation requirements, grantees may refer to the Evaluation FAQs located on the Knowledge Network.

An important consideration in selecting a data collection method is the available resources (funding, staff time, evaluation expertise) for carrying out the evaluation.
A program's budget for evaluation is always a consideration in selecting any form of data collection, as some methods can be more expensive than others. It's important to note, however, that virtually any form of data collection can be more or less expensive depending on several factors, including the length

or complexity of the data collection instrument or protocol, the number of respondents from which you plan to collect data, and whether data collection can be administered to groups of respondents at one time or have to be administered to individuals one at a time. Using existing data (which I will discuss in just a few slides), helps decrease the amount of new data that must be collected and by extension, evaluation costs.

In addition to considering the costs of data collection, it is also important to consider whether staff and/or volunteers have the time and expertise to assist in the data collection for the evaluation. Including staff in evaluation activities will not only help defray costs, but also help build evaluation capacity by involving staff in key evaluation processes.

When thinking about resources to invest in an evaluation, also consider whether outside expertise is required. It may be useful to consult an external evaluator to advise on data collection practices or have the evaluator design the data collection instrument (survey, interview protocols, etc.), while program staff may conduct the bulk of the actual data collection. Grantees that are required to conduct external evaluations, however, may need to negotiate this carefully to ensure objectivity.

Because most programs have limited resources that can be put towards an evaluation, it is important to remember that it is not necessary to evaluate every aspect of your program all at once as depicted in your logic model. Your evaluation can have a narrow focus (such as only addressing questions about ONE of your program's service activities and desired outcomes) or it can have a broader focus (such as addressing questions about EACH of your program's service activities and desired outcomes), depending on the information you hope to gain from your evaluation and the resources you have available.

In addition to considering available resources with respect to funding, staff time and external expertise, it is also important to identify whether data that you are currently collecting as part of routine program operations can be used for the evaluation. Grantees already engaged in performance measurement activities can build on that work as they plan for a program evaluation.

In the long-term, it is also important to consider how you can continue building on your data collection efforts for program evaluation over time. Think about building data collection into routine program operations from the start, if you have not already, rather than viewing it as one-time exercise for a specific evaluation only. Rather than a one-time activity, program evaluation should be thought of as part of a series of questions and answers that build upon one another over time that align with the life cycle of your program. This stepwise approach to evaluation is intended to generate more knowledge and evidence of your program's effectiveness over time.

We are now developing a webinar to describe a long term research strategy for you program but bottom line: collecting data regularly will pay dividends for years to come, for both performance measurement and evaluation and for continuous program improvement. Being a true learning organization means regularly collecting data and using it to inform decision-making.

There are four key questions to consider prior to selecting a data collection method for your evaluation. These include:

1. What is the purpose/objective of the evaluation?
2. What are the research questions?

3. What is the type of evaluation design?
4. What resources are available for the evaluation?

We will discuss each of these questions in the following slides.

In the second part of this presentation, we will review specific types of data collection methods and discuss their advantages and disadvantages.

Data can be obtained in many different ways to address your evaluation needs. Programs often begin by first looking at the data already being collected to understand if existing data can adequately answer the evaluation's research questions. Existing data (sometimes referred to as secondary data) may pertain to data already collected by the program itself such as performance measures or data that are gathered by external sources, such as administrative data. Administrative data may include student records, test scores, medical records, and Census data.

When existing data are not available to answer research questions, programs need to collect new data (sometimes referred to as primary data). Depending on the research questions being asked, there are several different types of data collection instruments that can be developed and administered to collect new data such as surveys and interviews.

There are advantages and disadvantages to both existing and new data. Existing data can save resources because time, effort, and financial resources are not spent on collecting new data. The disadvantage, however, is that existing data may be difficult to access, and there may not be a perfect fit between what the evaluation is trying to measure and the purposes for which the existing data were collected. Moreover, it is often difficult to assess how reliable the data are, and how reliably they were recorded and documented. Although more costly, the advantage of using new data is that the evaluator has greater control over the measures, procedures, and implementation and documentation, which can contribute to higher data quality.

As described earlier, grantees can build on data that are collected as part of routine program operations, such as performance measures, for their evaluation. Your program may already be collecting data to monitor and report on your program's progress toward achieving its expected outcomes for program beneficiaries. An example of a routine program activity from which you can gather data is the use of a sign-in log for any member trainings or workshops that are offered. You can use the sign-in log to track members' attendance in training to measure consistency of attendance as an indicator of members' commitment to development and learning. If your program decides to conduct an impact evaluation to learn whether your progress on outcomes is caused by your intervention, it may be that you continue to collect the SAME outcomes data, but also include a matched comparison group. Including this comparison group enables you to answer specific questions related to causality – in this case, what would have happened to people if they did not receive the intervention (i.e., whether the observed changes can be attributed to your intervention).

Besides thinking about existing and new data, there are two general types of data you can collect.

The first is Quantitative data. These data are numerical and can be counted, quantified, and mathematically analyzed. Examples of quantitative data are scores on achievement tests, number of program beneficiaries, or satisfaction ratings of AmeriCorps members. Surveys and assessments/tests are two of the most common ways of obtaining quantitative data. Because quantitative data are numeric, they can be easily ordered and summarized.

The second type is Qualitative data. These data are not easily reduced to numbers but rather narrative information that describes the study subject(s) and context. Some examples of qualitative data include transcripts of interviews and focus groups and field notes from observation of certain program activities. Qualitative data provide insight and context such as attitudes, characteristics, and perspectives of program stakeholders. Although there are many ways to collect qualitative data, interviews and focus groups are two of the most common ways of collecting these data.

Also, I want to highlight both quantitative and qualitative data are systematically collected, recorded, and analyzed. Thus, while individual anecdotes and testimonials may be useful for describing your program to outside stakeholders, they should only be used to highlight evaluation findings if they are systematically collected, recorded, and analyzed for the evaluation itself.

As we discussed, data collection methods can be either quantitative or qualitative. The table that you see here provides an overview of these two types of methods, highlighting their key differences with regard to the following:

Scope – In general, quantitative methods capture less in-depth data on a larger number of study participants while qualitative methods generate more in-depth data on fewer study participants. This is because quantitative methods are often used to collect data that can be summarized across a larger number of cases (e.g., study participants, program sites) whereas qualitative methods are often used to explore or gain a deeper understanding of a particular subject or topic among a smaller number of cases. Qualitative methods tend to be more labor intensive and time-consuming than quantitative methods approaches, and thus concern fewer cases.

Data collection – Quantitative data collection methods are used to collect data using structured instruments such as a survey with closed-ended questions or items (e.g., questions with pre-defined response options or rating scales). Qualitative data collection methods are intended to collect narrative data using instruments such as surveys or interview protocols that contain open-ended items (e.g., questions that allow the respondent to provide a response in their own words).

Data format – Quantitative methods yield numeric information, that is, data that can be summed up or counted. Qualitative methods, on the other hand, yield narrative data that provide important insights and context such as a description of the attitudes and perspectives of program beneficiaries.

Data analysis – Quantitative methods rely on statistical approaches to analyzing the data, which may range from generating simple statistics, such as frequencies, means, ranges, and standard deviations, to more complex statistical techniques that should be handled by an evaluator with expertise in

quantitative analysis. For qualitative data, content analysis is often used in which themes or patterns in the data are identified, categorized, coded, and then summarized.

Results – Quantitative methods produce results that can be generalized to a larger population. In program evaluation, generalizability may be defined as the extent to which you are able to make conclusions about a larger population based on information you have collected from a sample or subset of that population. For example, let's say the homelessness prevention program develops a new curriculum for its workshops and it first tests the curriculum on a sample of low-income families in the community that are assumed to be representative of other families in the community that would be served by the program. If this assumption is correct, the outcomes associated with using the new workshop curriculum on the small number of families can be generalized to estimate the outcomes of using the curriculum on other families in the community who did not participate in the workshop. Quantitative methods are generally the only way to provide statistical evidence of program impact. Qualitative methods produce results that give meaning, illustrative explanation, and views of study subject(s). Results are typically not generalizable to a larger population and cannot provide statistical evidence of program impact.

Most program evaluations will collect both quantitative data (numbers) and qualitative data (text) in a mixed methods design to produce a more complete understanding of a program – which is called triangulation. A combination of qualitative and quantitative data can improve a program evaluation by ensuring that the limitations of one type of data are balanced by the strengths of another.

Common quantitative data collection methods include surveys and assessments/tests.

Surveys are standardized instruments that collect data from a targeted group of subjects. While surveys may be used to collect qualitative types of data, on this slide we discuss their use for collecting quantitative types of data.

Surveys designed to capture quantitative data are generally comprised of well-specified, closed-ended questions with a limited number of open-ended questions for clarifying some topics. A closed-ended question is a question format that limits respondents with a list of answer choices from which they must choose to answer the question. An open-ended question is a question format that allows respondents to provide an answer in their own words.

Surveys can be classified by their method of data collection. The most common data collection methods are mail, e-mail/online, telephone interview, and in-person interview surveys. In choosing which method is most appropriate for your survey, it is important to consider factors such as cost and which method will likely yield the highest number of responses. For example, e-mail and online surveys are typically less costly; however, they would not yield a high response rate if your targeted survey group consists of individuals who are retired and over the age of 65. If response rates are low, you may be excluding valuable information that may result in an inaccurate account of the topic(s)/subject(s) your survey intended to measure. Thus, regardless of the method you choose, it is important to include ways to encourage participation in surveys, whenever possible. One effective strategy is to offer multiple

collection methods. For example, offer the survey electronically, or mail the survey with a stamped envelope. The easier it is for an individual to complete the survey, the more response rates will increase.

Assessments or tests are the use of established standards to assess knowledge, skill, or performance. They may be administered on paper, electronically, or via observation. Assessment tests are generally commercially available and have been independently validated for accuracy at measuring a particular concept, topic, or subject such as math or reading ability, parenting capacity, job readiness, etc. Programs may also choose to develop their own internal assessments/tests that are tailored to their program model. For example, an AmeriCorps program may work with their evaluator to develop an assessment that measures program participants' knowledge of a particular topic/subject(s) that their workshops focus on. The program may choose to administer this assessment at two time points - before program beneficiaries participate in the workshop series and immediately after they complete the series - to measure their beneficiaries' change in knowledge on the workshop's topic.

We now turn our attention to common qualitative data collection methods.

Interviews are a common form of data collection used to gather narrative data on individuals' thoughts, ideas, opinions, and perspectives, processes, and experiences. Qualitative interviews collect information by talking with and listening to people. Interviews can be completed either face to-face or over the telephone using unstructured, semi-structured, or structured questions to guide the discussion. It is common for interviews to involve audio-recording or a note-taker to ensure that respondents' responses are captured accurately and fully.

Focus Groups are another common qualitative data collection method. Focus groups involve collecting information through a guided small-group discussion. Focus groups are often used to collect information on topics that benefit from a collective discussion in order to understand the circumstances, behavior or opinions of focus group participants. When well executed, focus groups should create an environment that puts participants at ease, allowing them to thoughtfully answer questions in their own words and add meaning to their answers. The number of focus group participants may vary based on the questions you intend to ask, the available resources, and other factors.

Participant observation/field notes involve collecting information through observing program activities and study participants in their "natural" settings. There are different levels of "structure" that observations take, ranging from structured to unstructured. For example, in a highly structured observation, the evaluator will decide on the specific observation categories in advance. In a semi-structured observation, the researcher starts with an agenda of what will be observed and how, but collecting the data with observations is done in a less systematic or predetermined way. Unstructured observations would not be constrained by checklists and coding schemes; rather, the evaluator would report in narrative style about observations that are relevant to the research questions.

Regardless of whether the observations are structured or unstructured, it is generally advisable for a highly skilled qualitative researcher to conduct this type of data collection method. Also, since data obtained through participant observation often serves as a unbiased check against participants'

subjective reporting of what they believe and do, it is best to have this type of data collection conducted by an independent evaluator.

Document review involves the use of content analysis and other techniques to analyze and summarize printed material and existing information. Researchers often supplement other forms of data collection with the gathering and analyzing of existing documents produced in the course of everyday program activities. Because these materials are easy to obtain and do not take much effort to review, they can be a useful and inexpensive source of historical information on the program. The review of documents is an unobtrusive source of information that can provide a useful portrait of the program and/or program participants. Meeting minutes, program logs, announcements, training materials/manuals, policy statements, annual performance reports, etc. are all useful in developing a greater understanding of the program, its goals, strategies, activities, and participants.

I have covered a lot of information, I just want to pause here and see if you have any questions?

Before we continue, I want to ask you a few questions? Has anyone on the line tried to collect either quantitative or qualitative information or worked with grantees who have? Can you describe some of the advantages or disadvantages that you have personally experienced or have heard of others' experiences?

The next two slides highlight the advantages and disadvantages of each type of data collection method.

Starting with **Surveys**, the advantages of surveys for data collection are that they are:

- A quick and efficient way of getting information;
- Can cover a wide range of topics;
- Allow for the ability to obtain responses from a large number of people at once;
- Can be completed anonymously, if necessary, so respondents tend to be more comfortable completing them;
- Can be easy to compare and analyze responses using simple statistics such as frequencies, means, ranges, and standard deviations

On the other hand, surveys also:

- High response rates may also be difficult achieve but are important for ensuring quality data are used in analyses;
- Surveys that rely on closed-ended questions, where the respondent has to choose between provided response options, may lack in-depth information on a topic or subject. Analyses are then limited to response options that may not capture the nuances involved or the full range of experience on an issue;
- Can be unreliable if respondents misunderstand questions or choose not to answer and if not worded correctly, can produce biased responses;
- If face-to-face or telephone interviews are conducted, there can be the potential for respondents to answer questions a certain way in an attempt to please the interviewer. For example, if the

respondent is asked a question about their satisfaction with the program, they may not want to admit to the interviewer if they had an unsatisfactory experience.

The advantages of **Assessments/Tests** are that they:

- Provide objective information on knowledge and skills of participants;
- Because they tend to be standardized, assessment tests are easy to compare and analyze across respondents.

On the other hand, some disadvantages of assessment tests are that they:

- May be oversimplified and measure only a single, narrow construct;
- Also, assessment tests have been criticized for being biased against some subgroups of test takers, including low-income or minority populations.

Qualitative Interviews have several advantages, including:

- The ability to explore a range and depth of topics with respondents;
- Because of this ability to explore topics in-depth, interviews can yield rich data that can be highly useful and informative to programs;
- Qualitative interviews also provide an opportunity for the interviewer to explain or clarify questions that respondents may not completely understand.

On the other hand, the disadvantages of qualitative interviews are:

- They may be difficult to analyze and compare data depending on how structured the questions are;
- They usually require trained interviewers to complete them;
- Just as with face-to-face or telephone surveys, there is the potential desire of respondents to please the interviewer by answering questions in a way that they think the interviewer would agree.

The advantages of **Focus groups** include:

- The ability to efficiently obtain varying opinions and perspectives in a short period of time;
- Also, respondents can benefit from hearing the ideas provided by others in the group and building on these thoughts.

Disadvantages of Focus groups include that they:

- Require a skilled facilitator;
- It can be time consuming to analyze responses across multiple respondents and draw conclusions as this often involves transcribing audio recordings and coding transcript or interview notes to identify themes in the data.
- Finally, a common difficulty is scheduling a time and place to meet with many respondents at once.

The advantages of **Participant observation/ Field notes** include:

- The opportunity to view program operations in action;
- Participant observation provides direct information about behavior of individuals and groups that may contradict how individuals report and perceive their own behavior;
- Because the data collection occurs in a natural setting, there is generally little prep work that needs to occur prior to conducting this form of data collection. Observation also provides unbiased confirmation of participant behavior, unlike self-reported sources which are inherently biased because you are asking participants to report on their own actions.

Important disadvantages of Participant observation/ Field notes include:

- It generally can only be conducted by experienced and well-trained observers;
- It is always possible that the individual observer's selective perception may influence information collected;
- If the observers are not familiar enough with the setting, they may misinterpret certain behaviors observed. For example, they may observe teenage youth play-fighting with their friends and interpret their behavior as threatening to one another.

Some advantages of **Document Review** include:

- The information is generally easy to access;
- It can provide documentation of a program's history and development;
- Because documents may record occurrences over a long period of time, they can provide an opportunity to study historical trends, too.

Disadvantages of Document Review include:

- The reviewer is dependent upon the quality of the records, which may be incomplete or not available;
- Any analyses conducted would be limited to the data previously collected, so if the evaluator wanted additional information on a topic or area, another form of data collection would need to be used.

When identifying an appropriate data collection method for your evaluation, it is important to consider using more than one approach, resources permitting. A mixed methods approach, whereby a combination of qualitative and quantitative data methods are used, may yield more in-depth answers to your research questions. As we described earlier, each data collection method has its advantages and disadvantages and uncovers different aspects of the program. Using multiple methods provides an opportunity to triangulate your data – in other words, validate your findings through cross verification from two or more sources – and offers a more thorough assessment of your program.

Please see the handout of an example of using mulit-methods data collection for the AmeriCorps homelessness program:

When identifying an appropriate data collection method for your evaluation, it is important to consider using more than one approach, resources permitting. A mixed methods approach, whereby a

combination of qualitative and quantitative data methods are used, may yield more in-depth answers to your research questions. As we described earlier, each data collection method has its advantages and disadvantages and uncovers different aspects of the program. Using multiple methods provides an opportunity to triangulate your data – in other words, validate your findings through cross verification from two or more sources – and offers a more thorough assessment of your program.

Please see the handout of an example of using mulit-methods data collection for the AmeriCorps homelessness program:

Now that we have reviewed each of the data collection methods, in the next part of the presentation, we will review considerations in choosing a data collection method.

In addition to considering the advantages and disadvantages associated with the various types of data collection methods, there are other issues to consider in selecting the right data collection method or methods for your evaluation.

Research Ethics–Just as ethics in general refers to rules of right and wrong, there are ethics in conducting research. Research ethics refers to respecting the rights of human subjects, including their right to privacy, confidentiality, and respectful treatment. The foundation of research ethics is based on trust– study participants trust that the individuals collecting data are being honest with them about the purpose of and use for the data they are collecting; study participants also trust that researchers will keep their responses confidential; and public trust that the results reported by the program and evaluators are sound.

Institutional Review Board (IRB)-- Many different associations, government agencies, and universities have specific codes, rules, and policies relating to research ethics. The body that oversees and monitors these processes is known as an institutional review board (IRB). An IRB is a committee that has been formally designated to approve, monitor, and review research involving human subjects. It is important for you to be aware of any IRBs associated with your organization and/or your funders and what requirements they may have for approving and monitoring your evaluation activities. CNCS encourages grantees to submit their study research design and data collection methods to an IRB for review and approval.

Data Use Agreements-- Data Use Agreements are contractual documents used for the transfer of non-public data that is subject to some restriction on its use. This is important to keep in mind if you intend to use administrative data for your evaluation. When obtaining confidential data, many associations, government agencies, or universities require the individuals who will have access to the data to sign an agreement prior to its transfer. These agreements outline the terms and conditions of the transfer and address important issues such as limitations on use of the data, obligations to safeguard the data, liability for harm arising from the use of the data, publication, and privacy rights that are associated with transfers of confidential or protected data.

Given that many of you might be working towards outcome evaluations, I just want to review some additional considerations unique to outcome evaluations. I have two slides that may not be relevant for everyone, but please stick with me – I am getting near the end of the webinar.

Reliability-- is the ability to produce consistent results and indicates whether a study is likely to produce the same results if conducted again. While evaluation studies can be challenging to reproduce, there are many methods available to determine and improve the reliability of evaluations. For example, the reliability of a survey instrument can be explored by asking the same group of respondents the same or similar questions at two different times under similar conditions and see if they provide the same answers.

Validity-- The concept of validity is related, yet different from reliability. While reliability is about being able to reproduce the same results, validity is about coming as close to accurately measuring something as possible. For example, if you want to assess a child's math skills, the math test you administer to them must be validated such that it is measuring the child's actual math ability and not something else, such as their test taking skills.

The diagram shown demonstrates with dart boards the differences between reliability and validity. If the goal of darts is to hit the bulls eye on the board as often as possible, moving from left to right, the first dart board shows a situation where we have high reliability, but low validity. Reliability is demonstrated by the fact that the darts are consistently hitting the same area of the board over and over again. However, the darts never hit the bulls eye, indicating low validity or accuracy. An example of this situation might be a survey question that is worded in such a way that individuals consistently misunderstand the question and answer similarly, but incorrectly. The second dart board shows low reliability and validity because the darts are not consistently hitting any one area of the board and the darts never hit the bulls eye, which both indicate a lack of consistency and accuracy. An example of this situation might be a survey question that is overly vague and, therefore, individuals interpret the question in different ways and provide a wide variety of responses that are also inaccurate. Finally, the third dart board shows high reliability and validity because the darts are consistently hitting the bulls eye, demonstrating both consistency and accuracy. A well-worded question that is clear in what information it is requesting from respondents and is interpreted in a consistent way by respondents should have both high reliability and validity.

Sampling and generalizability– It is not always possible to include all potential program participants, otherwise known as a "population," in an evaluation. In these situations, studies often use statistical sampling to select a subset of individuals to participate in an evaluation who are representative of the larger population. Whether the findings of an evaluation are generalizable to the larger population is directly related to the representativeness of the sample used for the data collection. If done well, the results for the sample will reflect the results you would have gotten by surveying the entire group. Therefore, sample selection is extremely important to the usefulness of study findings. CNCS does NOT recommend using sampling for an evaluation if a grantee's sampling process will not result in a sample that is representative of the population they are studying. It is important to note that statistical

sampling is a complicated process involving a number of considerations, and, therefore, it often requires the assistance of a professional evaluator.

Statistical Power– An important consideration in determining the correct sample size is statistical power. Statistical power is one of several important considerations in determining the proper sample size for an evaluation. Statistical power helps determine how big a sample size is necessary to be able to detect a true effect of the program. In general (although it is not always the case), a larger sample size will provide more "power" to be able to detect a program's effect on its beneficiaries. However, it is usually not preferable to have a very large sample because of concerns about costs in terms of time, effort and other resources. So an evaluation generally wants to select a sample that is large enough to provide the necessary amount of power, while also meeting other considerations for completing the evaluation. A statistical power analysis is typically computed before conducting an evaluation to identify the optimal sample size that would be necessary to ensure the analysis is powerful enough to identify effects and also keeps the sample size as small as possible.

And lastly, covariates– Covariates are additional variables (other than the program outcomes being measured) that may affect evaluation findings. They are other factors that affect the relationship between the two main variables being studied (such as the program/intervention's effect on program outcomes). Covariates generally should be included in statistical analysis to eliminate the possibility that outcomes are due to other factors besides the program. A simple example is a study looking at the effect of a training program on math scores. In this case, it is reasonable to assume that there is going to be a lot of variation in the math ability of individuals enrolled in the training program– some individuals may have had more prior instruction in math than other individuals. If the evaluation does not control for these differences that existed prior to the program start, then it may be difficult to conclude if there is an actual effect of the training on people's math scores.

Now that we have reviewed data collection methods and their considerations, in this final part of the presentation, we provide examples of choosing data collection methods for different types of evaluation.

Before providing some examples of data sources and data collection methods for different evaluation designs, we first review and summarize the various data commonly available for an evaluation. Both process and outcome evaluations may draw on either existing data or new data to meet their evaluation needs, or both. Existing data include internal program records as well as external datasets. New data collection include surveys, assessments, interviews, focus groups, and observations/field notes. Because a process evaluation often seeks to explore questions about fidelity of implementation and program processes, this type of evaluation typically draws on several data sources capturing both qualitative and quantitative information. An outcome evaluation, on the other hand, asks about program outcomes and program effectiveness. For this type of evaluation, quantitative data collection methods are generally required.

This table presents an example data collection process for the fictional homelessness prevention program for low-income families and we've identified that this is for a process evaluation of the

program. The program is designed to prevent first-time homelessness in the county through a number of different activities. The main research question for the process evaluation concerns the series of workshops delivered by the program. Is the program's service activity – educational workshops - being implemented as designed?

Potential indicators for assessing fidelity to the program model include the duration of the workshops, and participant attendance rates which could be collected through member logs or records. For example, after each workshop, AmeriCorps members may be responsible for recording how long the workshop lasted, how many individuals attended the workshop, and what topics were covered during the session.

The evaluator can compile the data from all member's logs to assess whether the workshops are being implemented as designed and are consistent across members with regard to duration, attendance rate, and topics covered. Another potential indicator for assessing whether the workshops were implemented as designed are AmeriCorps members' delivery of the program curriculum.

This table presents an example data collection process for the fictional homelessness prevention program for low-income families and we've identified that this is for a process evaluation of the program. The program is designed to prevent first-time homelessness in the county through a number of different activities. The main research question for the process evaluation concerns the series of workshops delivered by the program. Is the program's service activity – educational workshops - being implemented as designed?

Potential indicators for assessing fidelity to the program model include the duration of the workshops, and participant attendance rates which could be collected through member logs or records. For example, after each workshop, AmeriCorps members may be responsible for recording how long the workshop lasted, how many individuals attended the workshop, and what topics were covered during the session.

The evaluator can compile the data from all member's logs to assess whether the workshops are being implemented as designed and are consistent across members with regard to duration, attendance rate, and topics covered. Another potential indicator for assessing whether the workshops were implemented as designed are AmeriCorps members' delivery of the program curriculum.

The evaluator may choose to use observations of the workshops on a quarterly basis to gather data on members' delivery of the curriculum. For example, the evaluator could develop an observation guide that lists the interactions, processes, or behaviors to be observed with space to record open-ended narrative data. The evaluator may focus on documenting interactions between the AmeriCorps members who are leading the workshops and the workshop participants, and on the AmeriCorps members' knowledge, skills, and behaviors. The evaluator will use the data to assess whether members are delivering the curriculum as intended and whether there is consistency across members in their delivery approach.

For continued discussion on reasons for selecting certain data collection methods for particular types of evaluation, grantees are encouraged to review the annotated sample evaluation plans that are posted on the Knowledge Network.

For this exercise, we ask that you identify appropriate data collection methods for an impact evaluation of the same fictional homelessness prevention program for low-income families.

The research question asks what impact the homelessness prevention program has on beneficiaries' ability to secure and maintain a stable housing status relative to a comparison group. The outcome of interest is the housing stability of low-income families at risk of homelessness.

Looking at the third column, what data can be collected to answer the research question and how?

[For facilitator: One way to collect this low income families' housing stability is through surveys. Another possibility is to identify any existing data that may address the research question.]

Moving to the fourth column, from whom would we collect this information?

For an impact evaluation, the information must be collected not only on program beneficiaries but also on an identified comparison group. Who might serve as a potential comparison group?

[For facilitator: In this example, the comparison group may be low-income families facing an imminent housing crisis who are receiving job assistance services through *another* program. Alternatively, programs may also look within their own program for a comparison group. For example, an impact evaluation may compare beneficiaries receiving core as well as supplemental services against participants receiving only core services.]

In the last column, we want to know when the data should be collected and by whom. What is a reasonable data collection timeline for this evaluation?

[For facilitator: Once the intervention and comparison groups have been identified, data will need to be collected at two time points. In this example, data will be collected both before the homelessness prevention program begins and a year after the program has been implemented for both the intervention and comparison groups. The evaluator may administer the survey or data may be collected by program staff, depending on the resources or particular arrangements made between the evaluator and the program.]

In the table above, we present one solution to answering the research question. You or your group may have identified a similar or different approach.

For this example, we can collect the appropriate information through surveys. The information, furthermore, must be collected not only on program beneficiaries but also on an identified comparison group for an impact evaluation. In this example, the comparison group pertains to low-income families facing an imminent housing crisis who are receiving job assistance services through another program. Alternatively, programs may also look within their own program for a comparison group. For example,

an impact evaluation may compare beneficiaries receiving core as well as supplemental services against participants receiving only core services.

Once the intervention and comparison groups have been identified, the evaluator will collect the data at two time points. In this example, data will be collected both before the homelessness prevention program begins and a year after the program has been implemented for both the intervention and comparison groups.

There are a few points that are important to remember when identifying the appropriate data collection methods for your evaluation:

The data collection method(s) that is most appropriate for your evaluation is one that most aligns with your evaluation's purpose, research questions, type of evaluation design and available resources. Also, keep in mind requirements. CNCS has different evaluation requirements for large and small recompeting grantees.

There are two general types of data collection methods – quantitative and qualitative – that can be used in any evaluation. Quantitative methods systematically document the existence or absence of program outcomes (such as knowledge, skills, behavior, condition or other occurrences), often by using surveys or assessments. These methods aim to yield unbiased results that can be generalized to a larger population. Qualitative methods, on the other hand, explore in detail the behavior of people and organizations and use observation, in-depth interviews, and focus groups and enrich quantitative findings. They help to understand the 'how and why' including explaining whether the program is likely to be the cause of any measured change. In cases where outcomes are not achieved, qualitative data can help understand whether this is a case of program failure or implementation failure.

There are different types of data collection methods that can be used in any evaluation. Each has its advantages and disadvantages. Therefore, no one approach is always best, and a carefully selected mixture is likely to provide the most useful information.

Lastly, process and outcome evaluations focus on different types of research questions and as a consequence, often demand different types of data collection methods. Because process evaluations tend to ask questions about implementation and fidelity to a program model, qualitative and quantitative data are typically necessary. Outcome evaluations, on the other hand, ask about changes and effects on beneficiaries and generally require quantitative data collection to answer.

Here we provide a list of resources on program evaluation and developing research questions that you may find helpful.