This presentation provides an overview of how to write an evaluation plan.

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

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

- Explain what an evaluation plan is and its purpose
- Identify the key sections to be outlined in an evaluation plan
- Determine what information to include within each section of an evaluation plan

The information contained in this presentation will help you work with an evaluator to develop and/or review an evaluation plan.

To facilitate your understanding of the information presented, we also have a few exercises that we've developed throughout the presentation.

An evaluation plan is:

- A written document that states the objectives of the evaluation, the research questions that will be answered, and how you plan to answer those questions. An evaluation plan in many ways is like a detailed instructional manual that captures the overall picture of what your evaluation will address and the evaluation steps and activities that are needed to successfully carry out the evaluation.
- It is important to understand that an evaluation plan should be thought of as a living document, such that it is continually updated as you make progress and finalize decisions about what your evaluation will entail. It is never too early to start your evaluation plan. While you may only be able to describe a few pieces of information at the beginning stages of the evaluation planning process, your plan should be a dynamic tool to help you establish your evaluation goals and objectives, organize your evaluation activities, and determine how the work will get done.
- Lastly, as you may already know, re-compete grantees are required to submit a draft evaluation plan as part of their AmeriCorps application. The evaluation plan that is submitted at the time of application should provide sufficient detail on the evaluation objectives and study methods, as well as plans for identifying a qualified evaluator and an estimated budget. Grantees are then expected to work with an experienced evaluator to further develop and refine their evaluation plan post-award. Thus, it is important for grantee staff to engage an experienced evaluator (either an external evaluator or someone internally who has evaluation experience) early on after receiving their AmeriCorps award to ensure that the evaluator is involved in developing the evaluation plan from the start.

This presentation covers what a well-written, fully developed evaluation plan should look like by the end of the first grant year of a re-compete AmeriCorps award.

There are many benefits to developing a written evaluation plan. Some of the key benefits are listed here on the slide. An evaluation plan:

- Helps the evaluation team be more systematic and comprehensive in their evaluation efforts
- Increases efficiency of the evaluation
- Helps anticipate any potential challenges upfront and avoid any missteps. For example, your evaluation plan should articulate potential challenges or limitations to the evaluation and strategies for addressing those challenges or limitations.
- Creates a shared understanding of the purpose and use of evaluation results for program stakeholders (e.g., funder, staff, other organizations, community members)
- Facilitates a smoother transition if staff turnover occurs
- Serves as a written understanding between the grantee and external evaluator(s), if applicable. While an evaluation plan is not a formal contract, it is an opportunity for the grantee to clarify their expectations.

Based on CNCS's guidelines in the Frequently Asked Questions: Evaluation document located in the Knowledge Network, when fully developed, your evaluation plan should include the following sections:

- Introduction
- Program background
- Research questions to be addressed in the study
- Evaluation design which includes the specific type of design that has been chosen for the evaluation
- Data collection
- Analysis plan
- Timeline
- Budget and other

Again, the outline presented on this slide can be found in the Frequently Asked Questions: Evaluation document, see FAQ 11, which is located in CNCS's Knowledge Network.

We will talk about each of these sections in more detail on the following slides.

The first section, the introduction, is typically a few paragraphs that help to orient your reader. The introduction is where you want to give a brief overview of what your program is and then talk about your goals for the evaluation, specifically what is it you want to learn or understand. The stated purpose of the evaluation drives the expectations and sets the boundaries for what the evaluation can and cannot deliver.

You also want to be clear about your evaluation's scope, that is, what aspect or aspects of your program your evaluation will focus on and over what time period. You want your introduction to explain who is sponsoring or requesting the evaluation and how and by whom the evaluation results will be used (e.g., who could make programmatic decisions based on use of the findings?).

Lastly, you want to introduce who your evaluation team is or who is going to be involved in the evaluation process, including each person's role and responsibility and their evaluation or research qualifications and experience.

The next section, the program description, will facilitate a shared understanding of the program between the program staff, the evaluation team, and any other stakeholders. Essentially, the work you do to lay out an accurate and comprehensive description of your program will be helpful in setting the stage for understanding your evaluation approach (e.g., selection of key research questions, evaluation design, data sources and methods, etc.). Your evaluation plan may be shared with individuals who are not as familiar with your program, so it is important to provide background information for them to fully understand how this evaluation fits in the context of the work you're doing.

Your program background section should include:

- Statement of need
- Your program's theory of change and supporting research evidence. This should be a synthesis of research that guides your choice of intervention and supports your theory of change, the cause and effect relationship you hope to achieve.
- Summary of previous evaluations that have been conducted on your program (if applicable)
- Narrative account of your program
- Your program's logic model (to be discussed in more detail on the next slide)
- Description of which program components the evaluation will focus on. After you've lay out a complete description of your program, it is important to reiterate to your reader which program components the evaluation will focus on. As you may know, it is not necessary to evaluate every aspect of your program as depicted in your logic model. Your evaluation can have a narrow focus (e.g., only address questions about one of your program's service activities and desired outcomes) or it can have a broader focus (e.g., address 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. This should be clearly stated up front in your evaluation plan.

Keep in mind that a lot of the information for this section can be drawn from existing program materials. For example, if you developed a detailed narrative and a logic model of your program for your AmeriCorps or other grant applications, you should be able to extract some of that information for your evaluation plan.

On the next slide, we provide an example of a program logic model for your reference.

On this slide we present an example of what a logic model might look like for a fictional AmeriCorps homelessness prevention program. This is on page 1 of your handout packet for easier viewing.

The following is a brief narrative description of the program (also on your handout): Increasing poverty and a growing shortage of affordable housing have contributed to an increase in first-time homelessness among families in a small rural county over the past decade. To combat the growing problem of family homelessness in the county, a program was created to meet the needs of low-income households facing a threat of homelessness due to a possible eviction or foreclosure. By providing targeted housing relocation and stabilization services (e.g., legal counsel, referrals to financial aid sources) and other assistance, the program is designed to address the housing crisis facing low-income families in order to prevent first-time homelessness.

The logic model we present here is a visual summary of the fictional affordable housing program. Logic models are typically read from left to right, employing an if-then sequence among key components. Based on the example shown here, it reads, if your program has these inputs or resources (listed in the first column), then it can carry out these activities (listed in the second column). If your program carries out these activities, then it can produce these outputs (listed in the third column). If your program has produced these outputs, then it will achieve these outcomes, ranging from short-term to long-term (listed in the latter three columns).

In addition, a logic model has two "sides." The **process** side focuses on how the program's carried out 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 hopes to accomplish, which can be short-term, medium-term, and/or long-term changes. The outcomes side reflects the changes, effects, or impact of the program. Essentially, what difference the program intends to make. ***Note:** It is not always feasible to measure long-term outcomes. Expected outcomes might be beyond the time frame of the program, or may be influenced by many factors outside of the program.

We want to note that logic models come in many sizes and shapes and also vary in level of detail, ranging from basic/simple to complex. There is no one or "right" way to develop a logic model. It often depends upon your purpose, how you will use the logic model, who will use the logic model, and what your program entails. The logic model we use here follows the CNCS template.

We will continue to refer back to this fictional program and logic model as we talk through other pieces of your evaluation plan.

The next section of your evaluation plan should articulate the key research questions that will be investigated. Research questions – the areas you'll examine as part of an evaluation of your program – are key to defining exactly what it is you're trying to accomplish. For that reason, research questions should be chosen carefully. When you choose research questions, you're really choosing a research problem - what you want to examine with your research.

There is no set number to how many research questions your evaluation should attempt to answer, however you want to be sure that it is feasible and you have the resources to answer each research question at the end of your evaluation. Thus, it is important that the research questions for your evaluation are stated in such a way that they:

- Are clearly stated
- Align with your program's theory of change and logic model. Your questions should match the objectives, activities, and outcomes of the program
- Will lead to measurable or observable results
- Are realistic to answer given your program's resources, time, and experience of those doing the evaluation. You want to make sure your evaluation team has experience evaluating similar programs or interventions and training that matches the type of design, methods, and/or approach of your planned evaluation.
- Builds upon results from previous research and evaluations (if applicable)

GROUP EXERCERCISE 1:

Let's now talk through a few examples of potential research questions for an evaluation. Please turn to page 2 of your handout packet where you'll find a list of potential research questions. I'd like us to go through each one as a group and discuss whether we think this is a strong or a weak question based on the first three criteria listed on the prior slide: clearly stated and specific, aligns with the homelessness prevention program model, and is measurable or observable. For this exercise, you can ignore the fourth and fifth criteria since those are program specific. For example, the fourth criteria will depend on an individual program's resource levels and experience of the chosen evaluation team.

Question 1: How did the low-income families in the county benefit from participation in the homelessness prevention program?

Question 2: What impact did the homelessness prevention program have on beneficiaries' knowledge of healthy food practices?

Question 3: Did program beneficiaries increase their awareness of housing, financial, and other types of services and benefits in the county as a result of their participation in the homelessness prevention program?

Question 4: Is the homelessness prevention program reaching its intended target population?

Question 5: Which family shelters and temporary housing have the best reputation?

Question 6: What do people in the county think about homelessness?

Answer guide:

- Questions 1, 3, and 4 are examples of strong research questions.
- Questions 2, 5, and 6 are examples of weak research questions:
    - Question 2 does not align with the program's theory of change and logic model because food security is not an intended outcome of the program.
    - Question 5 does not meet any of the three criteria. The question is vague, does not align with the program's theory of change and logic model, and would be difficult to measure or observe.
    - Question 6 does not meet the first two criteria – the question is vague and does not align with the program's theory of change and logic model.

As described previously, there are really two sides to your program which we showed in the logic model example. The **process** side focuses on how the program's activities are carried out or its planned work – inputs, activities, and outputs (direct products). The **outcomes** side of the logic model describes the expected sequence of changes that the program hopes to accomplish, which can be short-term, medium-term, and/or long-term changes. The outcomes side reflects the changes, effects, or impact of the program. Essentially, what difference the program intends to make. The stated objectives and goals of your evaluation go hand in hand with your evaluation research questions. If the objective of your evaluation is to answer questions about what you did in the program, the process side of the logic model will be where you focus your question(s). Likewise, if the objective of your evaluation is to answer questions about your program's outcomes or what changes the program brought about, the outcomes side of the logic model will be the focus of your question(s).

Process and outcome objectives generate different kinds of research questions. 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. It is important to note that a single evaluation plan can and often does include both process and outcome evaluation questions.

OPTIONAL SMALL GROUP EXERCISE #2:

Going back to the fictional homelessness prevention program, see page 1 of your handout, I'd like you to now work in small groups (based on your table groupings) to brainstorm and come up an example of a strong research question that a process evaluation of the homelessness program might address and an example of a strong research question that an outcome evaluation of the program might address. Again, remember that your questions should meet the three criteria presented on the prior slide: clearly stated and specific, aligning with your program model, and measurable or observable. Please come up with questions that are different from those presented earlier.

Facilitator – It may be helpful to turn back to the slide with the sample research questions (slide 10) so participants have visible examples as they work on this exercise. After 10 minutes or so, ask groups to volunteer to share the questions they developed. See the next two slides for examples of potential process and outcome evaluation questions which are option slides to present.

This exercise was intended to help you think through research questions for both a process evaluation and an outcome evaluation. To recap, a process evaluation provides information that helps you improve your program's operations and mainly focuses on inputs, activities, and outputs. An outcome evaluation provides information that can be used to demonstrate the results of your program on participants and the community. It focuses on the program's short-, medium-, and long-term outcomes.

On this slide, we have included a few examples of potential research questions for a process evaluation of the fictional homelessness prevention program. You may have come up with similar questions as a group.

Reach: Who did the homelessness prevention program reach in its first year of operation?

- What are the demographic characteristics of those served?
- To what extent did it reach its target population (i.e., low-income families facing an imminent threat of losing their home)?
- How many families did it reach overall and by service activity? Is the project reaching the intended number of participants?
- What recruitment strategies worked well in reaching the program's target population? What challenges did program staff face reaching the program's intended population?

Quality of implementation: How well was the program delivered?

- Was each service activity implemented properly, according to standards or protocol? Why or why not?
- What facilitated and what were barriers to the implementation of each activity? What lessons have been learned? How could the delivery of each activity be improved?

- How satisfied are the program's beneficiaries with the services they received?

Again, the answers to these kinds of questions allow you to assess whether program activities are occurring as you expected. They also can help determine areas in which a program needs improvement, so that you can reach expected outcomes.

Just to reiterate, these questions are examples of strong process evaluation research questions because they are: clearly stated and specific, align with the homelessness prevention program model, and are measurable or observable.

On this slide, we have examples of potential research questions for an outcome/impact evaluation of the fictional homelessness prevention program, focusing specifically on the overarching question of whether or not the program achieved its intended short-term results.

Potential research questions related to the short-term outcomes of the homelessness prevention program include:

- Did the families served by the program avoid losing their home due to an eviction or foreclosure?
- Did those served by the program increase their knowledge of responsible home ownership or tenant practices and skills?
    - How does their change in knowledge and awareness compare to a matched comparison group of individuals who did not participate in the program? (Note that this research question implies that your evaluation will assess the program's impact, that is, whether the program caused the observed changes in beneficiaries' knowledge)

Did those served by the program adopt more responsible home ownership or tenant practices and skills?

To recap, an outcome evaluation answers questions regarding program effectiveness; addresses whether a program is achieving its goals and objectives; and examines unintended consequences, both positive and negative. It provides information that can be used to demonstrate the results of your program on beneficiaries and the community.

And again, just to reiterate, these questions are examples of strong outcome evaluation research questions because they are: clearly stated and specific, align with the homelessness prevention program model, and are measurable or observable.

Up to this point in your evaluation plan, you've established the framework for your evaluation. After identifying the questions you want to answer, the next step is to explain the approach your evaluation will take to address those questions.

Evaluation design is essentially the structure that will provide you the information needed to answer each of your evaluation questions. The appropriate design for evaluating a program will largely depend upon certain considerations. This may include, but is not limited to:

- Your program's theory of change and logic model
- Primary purpose of the evaluation and key research questions

- Funder's evaluation requirements: Many funders require their grantees to complete evaluations of their program. As you consider different evaluation designs, you should also have a clear understanding of whether your funder has any specific evaluation requirements that must be fulfilled. Those requirements may drive the purpose and scope of your evaluation. For example, it is important to note that CNCS has evaluation requirements for large and small re-competing 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.
- Resources available for the evaluation: It is also important to take into consideration what resources (staff time, funding, evaluation expertise) are available to carry out the evaluation. In particular, your evaluation team must possess the skills and experience needed to carry out the type of evaluation design that is chosen. Think about whether you will be able to collect the data required for a given design. More complex evaluation designs require you to collect more data over multiple measurement points. This lengthens the data collection phase of the study and may require more sophisticated analyses to be conducted once the data are collected.

Your evaluation plan should clearly explain the evaluation design that will be used and a rationale for why that design was selected for the evaluation. Just as there are process-oriented research questions and outcome-oriented research questions, there are two common types of evaluation design that fall along those same lines:

- A process evaluation focuses on answering questions about your program's inputs, activities, and outputs. A process evaluation examines how well the program's implementation matches the theory behind its creation and confirms what the program actually does on the ground.
- An outcome evaluation addresses how a program's activities are related to changes, effects, and/or impacts on a program's stated outcomes, whether short-term, medium-term, or long-term outcomes. In general, an outcome evaluation measures program beneficiaries' changes in knowledge, attitudes, skills, or behaviors that are thought to result from the program.

Next, we are going to discuss each of these designs starting with the process evaluation.

We provide a few defining characteristics of process evaluations on this slide.

- A process evaluation focuses on a program's inputs, activities, and outputs
- It 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.
- The results of a process evaluation are most often used to change or improve the program. Process evaluations are able to address research questions about why a project is or is not successful, which can be very helpful for program staff and stakeholders because the results are useful for improving program practices.
- The collection of both qualitative and quantitative data through interviews, surveys, and program administrative data is usually preferred.

- Process evaluations mostly rely on simple descriptive statistics (means, frequencies, etc.) and do not require advanced statistical methods.
- To answer the types of research questions associated with a process evaluation, generally a comparison or control group is not necessary. We will explain what a comparison or control group is on the next slide as we transition to talking about outcome evaluation designs.
- It is also worth noting that the results of process evaluations are usually not generalizable, meaning that they can not be applied to similar program models being implemented in locations other than those participating in the evaluation. Note that a process evaluation design does not fulfill CNCS's evaluation requirements for large, re-compete AmeriCorps grantees (i.e., grantees receiving funding of $500,000 or more annually), but does for small, re-compete grantees. This is because process evaluations focus on measuring how the program is operating rather than measuring it's results (outcomes).

On the next few slides, we'll talk about a few different outcome evaluation designs.

A non-experimental outcome design only tracks or collects data on the outcomes for the intervention group (i.e., program beneficiaries). There are variations within the category of non-experimental outcome design, differing only in the number and the timing of outcome measurement points. These variations include:

- **Single group post-test**: Data are collected only once from study participants, immediately after they complete the intervention/program. The major drawback of any single group post-test design is that it does not provide a baseline against which results can be compared. To assess change, you must be able to compare post intervention data with a baseline measure. As shown in the first row of the table.
- **Single group pre- and post-test**: A pre- and post-test design measures outcomes among program beneficiaries before and after the intervention. With this design you simply administer the same measure twice, before and after the intervention. As shown in the second row of the table. The timing of the posttest measure is important. It should allow enough time for your program to have an effect, but not so much time that program effects are diluted or influenced by external factors such as participation in other programs or other environmental circumstances. You can improve upon the single group pre- and post-test design by additional measurement points post-intervention to gain a perspective on your program over time.

If you've chosen to use a non-experimental evaluation design, the number and timing of your outcome measurement points should be specified in your plan.

Compared to the other two outcome evaluation designs that we'll talk about next – experimental and quasi-experimental – a non-experimental outcome evaluation design does not provide the least credible evidence of program effects and therefore does not fulfill CNCS's impact evaluation requirement for large, re-compete grantees. Again, this is because a non-experimental outcome evaluation design does not follow a comparison or control group which is needed to produce estimates of what would have happened in the absence of the program. More rigorous outcome evaluations, such as quasi-experimental and experimental design studies, include a comparison or control group against which to

measure changes in program beneficiaries, and are thus able to measure or estimate the **impact** of a program.

A quasi-experimental evaluation design is distinct from a non-experimental outcome evaluation design in that it's defined by collecting data on two or more study groups – one is the intervention group (i.e., the program's beneficiaries) and the other is a comparison group (i.e. the study participants that do not receive the intervention or program services and thus serve as the basis for comparison when assessing the effects of an intervention on a program's beneficiaries). Including a comparison group enables you to answer specific questions related to causality – such as, what would have happened to people if they did not receive the intervention your program offers (i.e., whether the observed changes can be attributed to your intervention).

In a quasi-experimental evaluation design, the intervention and comparison groups are identified from pre-existing or self-selected groups and not through a random assignment process. We'll talk about random assignment when we turn to the next slide on experimental evaluation design. In a quasi-experimental design, for example, the intervention group often consists of individuals who self-select to participate in the program being evaluated while the comparison group consists of individuals who have similar characteristics to the intervention group EXCEPT that they are not served by the program or are participating in a different program. The key is that your comparison group does not receive the same intervention that is being evaluated because you want that group to serve as a point of comparison for the group that does receive the intervention.

It is important to note that you should be thoughtful in your selection of a comparison group to ensure that it is as similar as possible to intervention group. This is key to ensuring that evaluation findings are an accurate estimate of the program's effect. In most cases, you will want an experienced evaluator to use statistical matching to ensure that your comparison group is as similar as possible to the group of program beneficiaries in your evaluation.

Because the intervention and comparison groups are formed through a non random way, a limitation of quasi-experimental evaluation designs is that the study groups may differ in systematic ways, or in other words, in ways that are not due to chance, that then may account for some of the differences that are observed between the groups on the outcomes measured after the intervention. In a strong quasi-experimental design study, to reduce the potential for bias or error in the results, evaluators will attempt to use data (e.g., demographic data, baseline test scores) to make the intervention and comparison groups similar to one another on any observable characteristics that may affect the outcome(s) of interest. If using the homelessness prevention program as an example, the comparison group of families should be similar in size and structure (e.g., number of dependents, marital status), household income, parents' education attainment, and racial/ethnic composition. An experienced evaluator will be able to use advanced statistical methods such as propensity score matching or statistical controls to ensure that the treatment and intervention groups are comparable.

If you've chosen to use a quasi-experimental evaluation design, the number and timing of your outcome measurement points for both your intervention and comparison groups should be specified in your plan.

In sum, a quasi-experimental design provides comparatively stronger evidence of program impact than a non-experimental design, and if well-designed and executed, may fulfill CNCS's impact evaluation requirement for large, re-compete grantees. At the same time, a quasi-experimental design provides

comparatively weaker evidence of program impact than an experimental design. This is because pre-existing differences between the intervention and comparison groups at the outset of the intervention may lead to inaccurate estimates of the program's effects.

The image at the top right illustrates the most common form of a quasi-experimental design which includes a pretest and a posttest measure from both the intervention group and the comparison group.

An experimental design is similar to a quasi-experimental design in that it follows two or more study groups – one is the intervention group and the other is the control group. In an experimental evaluation design, the term control group rather than comparison group is used to describe the study group against which to measure the outcomes of program beneficiaries. The key difference between a quasi-experimental and an experimental evaluation design has to do with how the intervention and control groups are formed. In an experimental evaluation design, the evaluator employs random assignment techniques (e.g., lottery draw) to assign study participants to either the intervention group or the control group. This means that each study participant has an equal chance of being assigned to either the intervention group or the control group. The control group in an experimental design may receive no services or alternative or delayed services. The use of random assignment in experimental designs creates as equal groups as possible by ensuring that there are no systematic differences between the program and control groups. The process of randomly assigning program applicants for experimental design studies is complex and generally requires specific tailoring to each program's unique application and intake process. For this reason, random assignment is best conducted using a professional evaluator with experience completing these types of evaluations.

The key benefit of using random assignment is that it "evens the playing field." This means that the groups will differ only in the program or treatment to which they are assigned. Random assignment does not guarantee that the groups are "matched" or equivalent, only that any differences are due to chance. If both groups are equivalent except for the program or treatment that they receive, then any change that is observed after comparing information collected about individuals at the beginning of the study and again at the end of the study can be attributed to the program or treatment. This way, the evaluator has more confidence that any changes that might have occurred are due to the intervention that's being evaluated and not to the characteristics of the group. Evaluation studies that use an experimental design are often considered to be the most credible in regards to producing evidence of program impact. If well-designed and implemented, this design fulfills CNCS's impact evaluation requirement for large, re-compete grantees.

Again, if you've chosen to use an experimental evaluation design, the number and timing of your outcome measurement points for both the intervention and control group should be specified in your plan.

At this point, we've discussed the three main types of outcome evaluation designs: non-experimental, quasi-experimental, and experimental. As part of your description of the evaluation design that you've chosen to answer your research question(s), it is important to describe each of your evaluation study groups and how you plan to identify and recruit study participants.

Specifically, if you are conducting a non-experimental outcome evaluation design, only one group of study participants, that is, program beneficiaries, should be described. Quasi-experimental and experimental designs should describe both the intervention group and the comparison or control group

and include an explanation of what the comparison or control group will receive (e.g., no intervention, a different intervention).

To tie together our discussion of the three types of outcome evaluation designs, let's work together to complete this table by indicating which type of design best fits the description in the two columns on the right.

Facilitator – Ask the audience which of the three designs meet the description for each row. Then ask them which designs may fulfill CNCS's evaluation requirements for large, re-compete grantees.

Answer key: From bottom to top:

- Non-experimental
    - Using the homelessness prevention program as an example, a non-experimental outcome evaluation would only collect data on the outcomes of families who participated in the homelessness prevention program.
- Quasi-experimental
    - Using the homelessness prevention program as an example, a quasi-experimental evaluation would collect data on the outcomes of families who participated in the homelessness prevention program and a comparison group of families that did not participate in the program. The comparison group of families live in a neighboring county that does not have a homelessness prevention program and they are similar to the families that participated in the program with regard to certain demographic characteristics. Because the intervention and comparison groups were not formed through random assignment, the evaluator will use statistical techniques to make the two groups as similar as possible.
- Experimental –
    - Using the homelessness prevention program as an example, an experimental evaluation would collect data on the outcomes of families who participated in the homelessness prevention program and a control group of families that did not participate in the program. In this case, the program had more applicants for homelessness prevention assistance than could be served at any one time. Families would be informed at the time of program application that they will either receive the homelessness prevention program services or no services at all (at least while the evaluation is being conducted). The evaluator would then randomly assign the recruited study participants to either the intervention group or the control group. Because each family had an equal chance of being in the intervention or the control group, there should be no systematic differences between the groups, that is, no differences other than those due to chance, thus any differences that are observed in the outcomes can be attributed to the program.

The differences between national performance measure outcomes and evaluation outcomes are often confusing. If a grantee is planning to conduct an impact evaluation (e.g., quasi-experimental and experimental design studies), the addition of a control/comparison group provides an important element that differentiates evaluation from performance measurement. In impact evaluations, an

outcome is the change resulting from a program's activities or services relative to any change found in a control or comparison group.

For both impact evaluations and non-experimental outcomes evaluations, CNCS encourages grantees to build upon their national performance measure outcomes, when appropriate. However, they do not want grantees to simply report or reformulate these same outcomes for their evaluations. Instead, grantees should go beyond these measures in some way. Some options for grantees to consider include: 1) adding another outcome that is different from the performance measure outcome; 2) including one more post-test at a later point in time (6 or 12 month follow-up, for example); or 3) conducting an implementation evaluation alongside the outcome evaluation.

Let's now turn to talking about data collection. This fifth section of your evaluation plan is really about the data you will collect or compile to answer each of the key evaluation questions that you've identified for your evaluation, keeping within the framework of your chosen evaluation design. In this section, it is important that you explain what data will be used in the study, how the data will be used, and how you plan to collect or compile the data. You want to be sure that the data collection methods you're going that will help to generate accurate information about your program.

Key information to include in this section is listed here on the slide. We will talk through each of these components next.

Indicators (or variables) help define exactly what information will be used to answer your research question(s).

Data indicators will be either quantitative or qualitative (or a combination of both). *Quantitative data* is numerical and can be counted, quantified, and mathematically analyzed (e.g., GPAs, standardized test scores, attendance patterns). Surveys, questionnaires, and existing databases are the most common ways of obtaining quantitative data.

Qualitative data is non-numerical and used to provide meaning and understanding. Qualitative data capture information that is difficult to measure, count, or express in numerical terms. Though not useful for statistical analysis, qualitative data can provide important insights and context (such as attitudes and program stakeholder perspectives) that cannot be captured through quantitative data collection. Narratives of program beneficiaries describing their reasons for participating in your program are examples of qualitative data. Interviews and focus groups are two of the most common ways of collecting qualitative data, although other sources are possible.

If you are using an experimental or quasi-experimental design, the emphasis of your study will likely be on quantitative data, since these are the data to which statistical tests can be applied.

We provide a few examples of potential indicators that have been developed to answer two research questions for an evaluation of the fictional homelessness prevention program. You may refer back to the logic model on page 1 of your handout to refresh your memory on the program.

After you've described the activities and/or outcomes you want to measure and the indicators you will use to measure progress on them, you will need to discuss your selected data sources and data collection methods for gathering data on each of your indicators.

A list of potential data sources are presented here on the slide. The data sources you choose should align with your evaluation objectives, research questions and evaluation design. This process often begins by first looking at the data already being collected to understand if existing data can adequately answer the evaluation's research questions. Existing data may pertain to data collected by the program itself or data that are gathered by external sources, such as administrative or test score data. Data collection methods are the techniques used to gather the information needed to answer your research questions. When thinking about the method to use for collecting data, it is useful to consider which method is:

- More likely to secure the information needed;
- More appropriate given who is being asked to provide the information;
- Least disruptive to the program and target populations;
- Most feasible given the available resources

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 phenomenon 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., survey or interview questions with pre-defined response options) and rating scales. Qualitative data collection methods are intended to collect narrative data using semi- or unstructured instruments such as surveys or interview protocols with open-ended items (e.g., survey or interview 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 yield information that may be more difficult to measure, count, or express in numerical terms. For example, qualitative data may provide important insights and context (such as attitudes and perspectives of program beneficiaries in narrative form) that cannot be gained through quantitative data methods.

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 experienced evaluator. Qualitative methods rely on non-statistical approaches to analyzing the data, such as content analysis in which themes or patterns in the data are identified and coded.

Results – Quantitative methods tend to provide more objective results that can be generalizable to a larger population whereas qualitative methods tend to provide more subjective results that are typically

not generalizable 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. One last point about the difference between quantitative and qualitative results are that quantitative methods are generally the only way to provide evidence of program impact.

Lastly, 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.

Your plan should identify and describe your evaluation's timeframe for collecting information. This means that for each type of data you plan to collect, you should specify at what time point or time points you plan to collect the data. The timeframe you choose to collect your data should align with your chosen research questions and evaluation design. Your plan should also identify who will be responsible for collecting the data and your approach to managing and securing the data (i.e., outline how you will enter, track, store, and secure data).

It is often the case that the resources required to gather information from everyone or everything in your evaluation's target population of interest exceed the budget and available resources for the evaluation. Thus, the goal becomes finding a representative sample (or subset) of that population. Sampling is a statistically reliable way of identifying a representative group of persons from the entire population of your targeted study group. An experienced evaluator will be able to advise you as to whether or not you should select a sample for your evaluation and if so, the sample size you'll need, but it largely depends upon:

- Purpose of the evaluation
- Expected size of the study population,
- Data collection method that has been selected for the evaluation, and
- Evaluation budget and resources

You should review your options carefully and select the sampling technique that is most compatible with your evaluation purpose, design, and resources.

After data collection, the next section of your evaluation plan should describe the analytic techniques that will be used to analyze all of the data you plan to collect or compile for the evaluation. Analytical techniques range from basic counts to content analysis to complicated inferential statistical analysis. Depending on the type of data you plan to collect, you might need to conduct quantitative analysis, qualitative analysis, or both. The analytic techniques that are selected should be based on and align with your evaluation questions, the selected evaluation design, and the type of data that will be collected.

For quantitative data, any statistical analyses that you plan to conduct should be documented. While we do not attempt to go into detail on the different statistical techniques that might be used for quantitative analysis, most evaluations rely on simple descriptive statistics – means, frequencies, etc.

However, when more complex analyses and causal modeling are necessary (particularly in the case of impact evaluations), evaluators will need to use more sophisticated techniques such as analyses of variance, regression analysis, and so forth. If your evaluation design requires these types of complex analyses, it's best to consult an experienced evaluator to develop a plan for analysis. As mentioned before, quantitative analysis is generally the only way to provide evidence of program impact.

For qualitative data, your evaluation plan should describe the process that will be used to organize and aggregate the data into themes that allow you to identify data trends. To ensure that qualitative data (such as from interviews or field observations) are amenable to analysis and systematically comparable, coding schemes are typically applied to the notes or data. Various approaches are used to interpret meaning or themes from the content of text data. It is important to note that anecdotes or what we sometimes refer to as "great stories" are not the same as qualitative data. Anecdotes – or personal accounts, thoughts, or feelings – collected in an ad hoc fashion cannot tell us whether any improvements occurred in an intervention because no measurements were established.

Again, a mixed methods approach, whereby a combination of qualitative and quantitative data methods and analysis are used, may yield more in-depth answers to your research questions.

At this point, we've discussed three key sections of your evaluation plan – evaluation design, data collection, and analysis. What's presented on this slide is an example crosswalk that can be created to help you organize your data collection and analysis approach so that it aligns with each of your research questions. You may refer to page 3 of your handout.

The example presented here is based on 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.

Once the data have been gathered, simple descriptive statistics can be generated from the quantitative data such as frequencies on the use of the curriculum (e.g., which topics in the curriculum were covered

the most, least, not at all), average duration of each workshop, and average attendance rates. Meanwhile, qualitative data, such as observations of members' delivery of the program curriculum may be thematically coded and analyzed. Taken together, analyses of all the collected data are then used to assess the extent to which the program was implemented as designed.

On the next slide, we provide an example crosswalk developed for an impact evaluation of the program.

You may now refer to page 4 of your handout packet. In this example, we present a research question developed for an impact evaluation of the fictional homelessness prevention program. The 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. One way to collect this information is 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.

After the data have been gathered, statistical tests (in this case, difference-in-differences methods) are then used to compare program participants with their matched comparison group by subtracting the average outcome (gain) in the comparison group from the average outcome (gain) in the intervention group. Such analyses may show that, on average, low-income families facing homelessness participating in the program are more likely to secure stable housing and remain housed than low-income families facing homelessness that were NOT served by a homelessness prevention program. It is important to note that the statistical techniques and methods used in an impact evaluation (involving for example, propensity score matching, identification and inclusion of covariates, etc.) should be conducted by a qualified evaluator.

Let's turn now to one of the final sections of your evaluation which concerns your evaluation timeline. Your evaluation plan should include a detailed timeline of when you expect to carry out each of your key evaluation activities specified in your plan. What's presented on this slide is a much simpler example of an evaluation timeline over a three year period to give you a general idea of some of the key evaluation activities to consider as you plan for your evaluation. You may find it helpful to lay out a more detailed timeline that lists each of your key evaluation activities and maps out when you expect the activity to start and end and who is responsible for carrying out the activity for added accountability. You want to make sure your timeline is feasible and aligns with any requirements set forth by your funder.

An estimated budget for your evaluation also should be included in your plan. The cost of evaluations varies widely and will depend on the type of study design, the size of the study, the level of expertise and experience of the evaluator, and data collection expenses. Other common considerations for creating a program evaluation budget are: staff time; materials, equipment, and supplies; travel costs;

and data collection. With respect to this last item, evaluations involving more primary data collection tend to be more expensive than those that rely on existing internal program records or external data sources. This is not a comprehensive list of cost consideration. Depending on the program to be evaluated and/or the actual evaluation activities, there may be additional expenses required.

If there is additional information relevant to the evaluation that does not align with the other key sections of your plan, consider it in this section. One example would be information on obtaining Institutional Review Board (IRB) clearance, if that is deemed necessary for your evaluation. IRBs are committees that review research protocols and other materials to ensure the rights, safety, and welfare of human subjects participating in studies. The IRB can certify that the rights of subjects will be protected, that any potential adverse effects on participants will be minimized, and the data will be securely managed and maintained. IRB clearance may be required by your agency leadership or funding source, so it is important to understand what is required before you begin to recruit study participants and collect data. An experienced evaluator should be aware of any IRB clearance requirements that will need to be fulfilled before data collection can begin.

The appendix of your fully developed evaluation plan should include a list of references and copies of any data collection tools or instruments you intend to use for collecting data. Note that copies of data collection tools or instruments are not required elements of the evaluation plan that you submit during the Grantee Application Review Process (GARP) but should be included in your fully developed evaluation plan by the end of the first grant year.

GROUP EXERCISE

Now that we've walked through each of the key sections of your evaluation plan, we have one final group exercise that we'd like you to participate in. Please turn to page 5 of your handout. You'll find an example written evaluation plan for a fictional AmeriCorps program. We ask that you read through the sample plan and then critique the plan in small groups based on what you've learned are the key components of an evaluation plan. We have provided a checklist as a guide for reviewing the plan and its components which is the second handout you should have received (single sheet only).

Facilitator – Ask the audience to share their thoughts on the sample evaluation plan. Walk through each section and ask for input on strengths and weaknesses, what might be missing, how could the section be improved upon, etc.

As we conclude today's presentation, we've highlighted some things to keep in mind as you work to develop and use your evaluation plan:

- Developing an evaluation plan is a collaborative process. This means that you want to involve a number of different stakeholders (e.g. program staff, the evaluation team, and other stakeholders) to ensure that the findings from your evaluation will be useful.
- An evaluation plan is a dynamic tool and can change and be refined as you make decisions about how best to evaluate your program. It's important to remember that your evaluation plan is a living document that you can continue to develop and refine as your evaluation and/or program evolves.
- An evaluation plan facilitates the process of keeping diverse stakeholders on the same page with regards to the actual implementation of the evaluation. Because a number of different

stakeholders have given input into your evaluation plan, the actual implementation of your evaluation should go more smoothly.