



DEFENSE CENTERS OF EXCELLENCE

For Psychological Health & Traumatic Brain Injury

Analyzing Program Evaluation Data: How to Interpret Quantitative Data on Processes and Outcomes

Presented on May 19, 2015

Episode 6 in the FY2015 Program Evaluation and Improvement Training Series

Presenters:

Capt. Armen Thoumaian, Ph.D., USPHS
Health Science Officer
Office of Shared Services Support
DCoE

Richard Best, Ph.D.
Research Scientist
Contract support for DCoE

Carter Frank, M.A., M.S.
Research Scientist
Contract support for DCoE

Moderator:

Carmina Aguirre, M.A.
Research Scientist
Contract support for DCoE

[Video Introduction]

Capt. Thoumaian: Hello. My name is Captain Armen Thoumaian of the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury or DCoE. Thank you for joining us for this episode of the DCoE Program Evaluation and Improvement webinar training series.

DCoE's Mission is to improve the lives of our nation's service members, families and veterans by advancing excellence in psychological health and traumatic brain injury prevention and care.

DCoE accomplishes that mission in coordination with its three Centers: Defense and Veterans Brain Injury Center, Deployment Health Clinical Center and National Center for Telehealth and Technology. Together, we produce a variety of trainings on subjects ranging from program evaluation to clinical care and prevention practices.

This training series is designed for program administrators and service leadership who are involved with or who plan to conduct program evaluation activities within the Defense Department's psychological health and traumatic brain injury programs. Our objective is to enhance the capability of these personnel to actively engage in program evaluation activities

and, ultimately, make program evaluation an inherent component of everyday program operations.

By supporting enhanced program evaluation capabilities across the Defense Department, this series contributes to DCoE's larger mission to improve the quality and effectiveness of the psychological health and traumatic brain injury prevention and care programs that serve our military members, their families and veterans.

On behalf of DCoE, thank you for participating in this training series.

[Slide 1]

Ms. Aguirre: Hello. My name is Carmina Aguirre. I provide contract support to the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury or DCoE. I will be your moderator for this presentation, the sixth episode in the 2015 DCoE Program Evaluation and Improvement webinar training series. The webinar is hosted using the Adobe Connect platform and the technical features are being handled by DCoE's webinar support team in Washington, D.C.

Today's topic is "Analyzing Program Evaluation Data: How to Interpret Quantitative Data on Processes and Outcomes." Before we begin, let's review some details.

[Slide 2]

This presentation has been pre-recorded; however, there will be a live Question-and-Answer session at the end of the presentation.

Throughout the webinar, we encourage you to submit technical or content-related questions using the Question pod on your screen. Your questions will remain anonymous, and our presenters will respond to as many questions as possible during the Q-and-A.

At the bottom of the screen is the Chat pod. Please feel free to identify yourself to other attendees and to communicate with one another. Time is allotted at the end of the presentation to use the Chat pod for networking.

All audio is provided through the Adobe Connect platform; there is no separate audio dial-in line. Please note there may be delays at times as the connection catches up with the audio. Depending on your network security settings, there may also be some noticeable buffering delays.

Closed captioning is provided for today's event, and a transcript will be made available at a later date.

[Slide 3]

Webinar materials for this series are available in the Files pod at the bottom of the screen during the webinar. They are also posted in the Program Evaluation section of the DCoE website. Modules from the newly revised DCoE Program Evaluation Guide will be posted throughout the 2015 webinar series.

For information about other DCoE webinars and trainings, visit the Training section of the DCoE

website by following the link on slide 3.

[Slide 4]

We are pleased to offer continuing education credit for the 2015 Program Evaluation and Improvement webinar series. Instructions for obtaining continuing education through DCoE's collaboration with the Professional Education Services Group were made available during the registration process. Eligibility criteria for continuing education credit are presented on slide 4. In an effort to enhance the focus of individual webinar episodes, we have reduced the length of this and future episodes to one hour. As a consequence, please note that eligible participants will receive one hour of credit rather than an hour-and-a-half.

[Slide 5]

If you preregistered for the webinar and want to obtain CE certificates or a certificate of attendance, you must complete the online CE evaluation. After the webinar, please visit dcoe.cds.pesgce.com to complete the online CE evaluation and download your CE certificate or certificate of attendance. The CE evaluation will be open through May 26th, 2015.

[Slide 6]

This webinar was introduced by Captain Armen Thoumaian. Captain Thoumaian is the Deputy Chief of the Office of Integrated Services at DCoE. He is a Scientist Director in the Commissioned Corps of the U.S. Public Health Service with more than 30 years of experience in health and mental health program design and evaluation. In January 2012, Captain Thoumaian joined DCoE to help design and implement program evaluation and improvement efforts in the Defense Department. He holds a B.A. in psychology and sociology, an M.A. in general experimental psychology, and a Ph.D. in social welfare and social work. Captain Thoumaian has also completed a National Institute of Mental Health fellowship in Community Mental Health.

[Slide 7]

Presenters for this episode include Dr. Richard Best and Mr. Carter Frank. Dr. Best is an industrial and organizational psychologist with 14 years of experience conducting health services research in both the Veterans Health Administration and the Defense Department's Military Health System. He has extensive experience in research design, qualitative and quantitative data collection and analysis, and collaborating with clinical experts to translate research results into actionable recommendations. Dr. Best holds a master of science and a Ph.D. in industrial-organizational psychology and is certified in Prosci's change management process.

Our other presenter is Mr. Carter Frank. Mr. Frank is also a research scientist who provides contract support to DCoE. Mr. Frank has over 15 years of experience in program development and management at local, regional and national levels. The breadth of his 33-year career includes 11 years of military service, spans military and civilian environments, clinical and non-clinical mental health operations, training, human resource management, business development and government contracting. Mr. Frank holds a B.S. in mathematical sciences and master's degrees in counseling and management information systems. He is a licensed clinical professional counselor.

[Slide 8]

I am Carmina Aguirre, your moderator for today. I am also a research scientist who provides contract support to DCoE. I have over 14 years of experience within the Defense Department. My background includes executive leadership, psychological health, sexual assault prevention and response and public affairs. In addition to supporting DCoE, I serve as Chief of Public Affairs in the Florida Air National Guard. I hold a B.A. in psychology and an M.A. in human services with a specialization in executive leadership.

[Slide 9]

This training presentation will provide additional depth and guidance on analyzing and interpreting quantitative data on processes and outcomes for program managers and others involved with program evaluations. Topics will include an introduction to quantitative data analyses, general strategies and concepts integral to processes and outcomes, and consideration of common challenges that arise when analyzing and interpreting data.

At the conclusion of this webinar, participants will be able to:

- Explain key types of quantitative analysis used in program evaluation
- Demonstrate basic knowledge of how to use program data to examine process and outcome metrics
- Perform basic quantitative analyses and interpret quantitative findings
- Select and implement strategies to address common challenges related to quantitative data analysis.

[Slide 10]

As seen on slide 10, Captain Thoumaian will begin with an introduction to quantitative data analyses. Dr. Best will then discuss how to analyze and interpret program processes, followed by Mr. Frank who will review how to analyze and interpret program outcomes before presenting strategies for overcoming common challenges. We will then conclude with a summary by Captain Thoumaian. Then, I will provide a list of references and resources, followed by an opportunity to provide anonymous feedback and a brief question-and-answer session with our presenters.

[Slide 11]

Thank you Ms. Aguirre. Quantitative methods are an important source of information for program evaluation efforts. The content in this portion of the presentation is intended to apply to a wide range of psychological health and traumatic brain injury program managers. The content will “build” and become very specific to program evaluation. For now, we begin with a general introduction to analysis techniques to help make the evaluation data you collect more understandable. This portion of the presentation provides a basic overview of the most commonly-used analysis techniques.

[Slide 12]

Stephen Few, an author and business analyst, said, “Numbers have an important story to tell. They rely on you to give them a voice.” The focus of this training is about making sense of the data we collect and learning what they have to tell us. It is about what to do with data once they have been collected.

You don't need to have extensive training in order to have a basic knowledge of how to interpret findings. That said, this introduction is not intended to be a substitute for a solid research methods course. It is meant to be a quick guide to key concepts and information. We also provide references, resources, and reading recommendations at the end of the presentation on slides 53 through 56, for those who want to learn more.

[Slide 13]

Data analysis enables us to describe large amounts of information and discover patterns in the data. In other words, analysis provides the voice that allows the data to tell their story. In order to promote high standards of prevention and care among psychological health and traumatic brain injury programs, we need to focus on interventions and practices that have the greatest evidence of effectiveness. We need to understand how well a program is working. This requires collecting information, analyzing data, and interpreting what the data mean.

[Slide 14]

The first step in analysis is to convert data accurately into a clean, usable form. Deciding what to do with "missing" or "not applicable" items should be determined in advance. These decisions should be documented for future reference. For example, a person may leave a survey answer blank because they don't know the information, or the question may not actually apply to them. Leaving blanks will interfere with higher-level analyses.

Before actually conducting analyses of the data, some editing checks should be performed to reduce the chance for error. Coding mistakes are a major source of error in survey measurement. For example, an answer of 5 on one question might mean that it was rated "the best," while an answer of 5 on another question might mean it was "the worst." In this example, the ratings of 1 through 5 can have opposite meanings for different sections of the survey. Thus, the data will need to be re-coded so that all answers are interpreted in the same direction and can be summarized and considered together.

When responses to an item are so few that analyzing and reporting on it becomes meaningless, it may need to be combined with other categories, and the category title or label changed accordingly. Generally, there should be 5 or more responses to a question to avoid what is known as "the small cell problem" to allow for greater diversity in responses.

Look at summary results when analyzing the items on your measures to be sure they make sense. If you have 400 program participants who are supposed to answer an item that asks "What is your gender," the summary results should NOT total to 425 participants. As you review your data, you might ask yourself if the responses are realistic.

[Slide 15]

The aim of data analysis and interpretation is to condense large amounts of information into a usable form and communicate major findings. Ask the wrong person how their weekend was, and you may be treated to a detailed chronology of every minute rather than a summary description. Data analysis should provide enough detail to be informative but not so much that the audience cannot understand and absorb it.

Generally, there are two types of data analysis: descriptive and inferential. Descriptive data

analysis describes the data and can help you learn what is average or typical for your program as well help you understand the spread or variation among participants, processes, costs, and outcomes.

As we will discuss in slides 16 through 19, descriptive analyses include a variety of techniques such as frequency counts, measures of central tendency, and measures of variability.

Choose the descriptive analyses that best answer your evaluation questions. If a question asks about the number of pamphlets produced or the number of sessions delivered, a frequency count may suit your needs.

The measures of central tendency, such as the mean, median or mode, as well as measures of variability, such as the range or standard deviation, will help you inspect your data for errors, missing or incomplete data, or outliers. These measures will also help you determine whether or not your data are normally distributed, which is important if you wish to also use inferential analyses.

[Slide 16]

Numerical counts, or frequencies, tell us how many times something happened, or how many responses fall into a particular category. For example, you can say that

- 82 participants are over 25 years old, or that
- 105 of the 130 participants said program sessions are very useful and help improve family communication problems.

Frequency counts can also serve as a base for other calculations, such as for percentages.

When working with percentages, it is common to include the count such as:

- 105 participants among the total number of 130 who responded, or 81 percent, rated the program as very useful.

Reporting percentages is a very useful way to think about numerical data, so there are a few rules you should follow.

First, use the correct denominator, or base from which the percentage is calculated, and communicate to others which base you are using. For example, does 75% mean:

- 75% of all participants,
- 75% of those who completed a survey,
- 75% of those who answered a specific question, or
- 75% of those to whom that question applied?

When a large number of responses are “missing” or marked as “not applicable,” the correct base may be the number of participants who answered the question, rather than the total number of participants. Using the total number of participants as the base could result in an incorrect percentage. When reporting percentages, such as answers to a survey, show how many “missing” responses there were.

Second, round up to one decimal place even though many software programs provide results out to several more decimal places. The general guidance is to round numbers 5 or greater up to the next number, as shown on slide 16. Here, you would report the score as 17.6 rather than 17.5714 to make it easier for the audience to understand.

Adding percentages can be tricky: If you have a questionnaire with a “check all that apply” response category, your percentages will add up to more than 100%. You can only add percent responses when the answer categories are mutually exclusive.

Lastly, when computing an average percentage, always use the original numbers. Do NOT sum several percentages and divide by the number of percentages that you added together. This can allow errors to creep in.

[Slide 17]

Measures of central tendency describe what is typical for a distribution of scores or for a group. The most commonly-used measures of central tendency are the mean, median and mode. The mean is what we think of as the average; the sum of all values or scores divided by the total number of participants. The median is the mid-point or middle value; half of the values fall above the median and half fall below. The mode is the most frequently occurring value in the distribution of scores or values.

[Slide 18]

In addition to measuring the central tendency of a distribution of values, it is also common to report the variability: the dispersion, or spread of the distribution. The range expresses the distance between the lowest and highest scores. On slide 18, bar graphs for programs 1 and 2 feature the same mean, but you can see that participant scores are more widely dispersed for Program 2 than for those of Program 1. The range informs us of the lowest and highest scores, but does NOT tell us anything about the distribution of the scores, which is represented in the standard deviation.

[Slide 19]

The standard deviation provides information about how far, on average, the data move away, or deviate, from the mean. It makes use of every score in the distribution, as we see represented in the center of slide 19, also known as “the Bell Curve,” or normal distribution. This is a naturally-occurring phenomenon that has been well-observed for many measured variables. We see that most scores cluster around the mean, with some tapering off into “tails” at either end.

On the bottom left is a distribution in which most responses are similar to the mean indicating small deviation or a low standard deviation, while on the upper right, we see that most responses deviate greatly from the mean illustrating a large standard deviation. When all responses are identical, the standard deviation is zero.

Sometimes, variation represents a positive outcome: A program designed to help people think independently and build individual decision-making skills may reveal a variety of perspectives. If the goal of a program is to help everyone achieve a certain level of knowledge or skill, however, large variation may not be the desired outcome.

[Slide 20]

Inferential statistics help tell the story by allowing you to make inferences based on the results of the analyses. They empirically assess whether a change is meaningful or significant. For example, you can use inferential statistics to determine whether a change in program participants’ knowledge is significantly related to the activities and products produced by the

program.

In addition, inferential statistics enable you to generalize or infer findings from one group to a larger population and to see whether that relationship is truly meaningful. Relationships found in one group will not necessarily hold true for the wider population. Inferential statistics assess whether a relationship between variables is negative or positive, the strength of that relationship, the likelihood of that relationship occurring again, or whether that finding was simply due to chance. This type of analysis will require some expertise, possibly involving outside resources.

Inferential statistics enable inferences about the data, such as whether one or more variables are associated with a particular outcome variable, or that findings from one sample can be generalized to a larger population. These concepts will only be briefly described.

[Slide 21]

A comparison or control group can help determine whether an intervention, such as a program activity, specifically caused changes in participants. For example, if you wish to evaluate whether your program reduces depression in your participants, a comparison group can help in that analysis. A comparison group is similar to the experimental or intervention group in characteristics such as age, gender, rank, ethnicity, service branch, and so on. The main difference between the comparison and intervention groups is the fact that the comparison group did NOT receive the intervention or did NOT participate in the program. By comparing these two groups on an outcome of interest, for example depression scores, you can better tell whether the program had the intended effect, such as reducing depression.

Again, some statistical expertise will be required to establish that the groups vary only in whether they receive the 'intervention'.

[Slide 22]

Check the data for patterns. The answers to some questions may seem to link with responses to other questions. Displaying the results of your data analyses in pie charts, bar graphs, tables, lists and line graphs will help reveal possible patterns. See how the data look when displayed differently. Does anything 'jump' out at you? Using charts and graphs also may help you organize your report-writing and communicate your findings more effectively to others.

Crosstabs, a form of sorting the data into meaningful categories, will allow you to get an overall picture of what the data are telling you. For example, do satisfaction ratings from program participants vary by age, rank or education? Crosstabs allow you to examine patterns among subgroup categories. Examining patterns at a deeper level can help you determine what you may want to investigate further using inferential statistics.

[Slide 23]

Before you begin data analysis, be sure that you have the materials and resources required for this task. A few are listed here, on slide #23. Also consult your leadership and similar programs to see what resources are available and which ones are needed.

And now, Dr. Best will discuss how to analyze and interpret program processes.

[Slide 24]

Thank you, Captain Thoumaian. The purpose of this section is to provide an overview of analyzing and interpreting program processes as part of an evaluation effort.

[Slide 25]

To get started, it helps to understand various components of process analysis and the program, such as:

Resources – the facilities, staffing, space, financial and other physical resources that were needed at program implementation and that are needed currently in order to operate.

Barriers – did the program encounter any barriers during the planning and implementation stages and how did those barriers affect the program? Barriers can include inadequate funding at implementation, reduced funding during program operation, lack of training for program staff, or not having the number or type of staff needed to implement and run the program.

Services/activities – what sort of clinical, outreach, or educational activities are offered by the program to achieve the desired effect on program participants or the target population? Are these activities in line with the program's mission statement, goals and objectives? Are they linked with outputs and outcomes in the program's logic model?

Exposure – who has access to, or is exposed to the program? How are these individuals made aware of the program? Does the program have a strategy to recruit participants? Does it have a way to keep participants in the program? In short, does the program track or follow participants to ensure they are receiving the services they need?

Finally, **Context**, includes the environment in which the program operates. For example, environmental factors, such as service mandates or directives, or changes in mission may impact the program and thus cause the program to change over time.

[Slide 26]

As seen on slide 26, there are three categories of processes with several examples of metrics that fall within each one.

The first process category, **Participation**, refers to tracking metrics such as the number of calls to a helpline, session attendance, the number of sessions held, the extent of the target population reached, participant demographics (for example; rank, branch of service, sex) and perhaps how the participant was referred to the program.

A second process category you can track is **Program Satisfaction**. This process category allows you to track satisfaction and feedback from important stakeholders like participants, staff, and key leadership. You can track satisfaction ratings using quantitative measures such as questionnaires or evaluation cards or qualitative methods such as interviews or focus groups. These methods can also be used to track other important information such as how likely a person is to refer other individuals to the program.

Finally, **Activities** can be tracked by such metrics as the frequency and length of each program activity, the number and type of program activities, or the number of referrals generated.

In the next several slides, we provide a sample evaluation question that is appropriate for process analysis, followed by examples of process analyses for a hypothetical program, including metrics for the target population's demographic characteristics, participant satisfaction ratings, and the frequency of program activities.

[Slide 27]

Slide 27 provides an example of an evaluation question that is appropriate for a process analysis: "Was the program implemented with fidelity?" This is an important question to answer as programs generally are designed on the basis of scientific evidence and have structured processes that intend to change program participants in some meaningful way. Thus, it is important to ensure the program is implemented with the same evidence-based processes. Other questions of interest may include: How does the program operate? What is the program expected to achieve? How is the program expected to accomplish what it has set out to achieve? and finally, How do participants perceive the program?

The purpose of asking such questions is to not only analyze program processes but to answer the question; "How do these processes affect program outcomes?" Answering the question about how participants perceived the program, for example, provides useful information for understanding how and why the program did or did not change participants.

The linkage between program processes and program outcomes facilitates the measurement of outcomes, which will be discussed in the next section of the presentation.

[Slide 28]

To better illustrate Process Analysis, we refer to a hypothetical reintegration program called Program Sierra, formerly referred to as Program Echo in previous webinars and Program Evaluation Guide modules. Program Sierra's¹ mission is to ensure that service members who are wounded, ill or injured successfully reintegrate into civilian life or return to active-duty in the military. As part of this mission, Program Sierra personnel hope to enhance force readiness and improve the quality and efficiency of services across the Defense Department.

For convenient reference, we have also provided Program Sierra's mission, goals, SMART objectives and logic model in the extra slides at the end of this presentation.

[Slide29]

Program Sierra's evaluation team, led by its program manager, decided to conduct a process evaluation to determine whether the program was implemented with fidelity. As part of the evaluation of Program Sierra's fidelity, key stakeholders and leadership were also interested in:

- determining the percent of the target population being reached,
- the demographic characteristics of the program participants, and
- the satisfaction of the program participants

Answers to these questions can help the team of evaluators determine program improvements that can enhance the quality and effectiveness of Program Sierra.

¹ Program Sierra was formerly known as Program Echo.

[Slide 30]

There are various metrics that can be used to analyze Program Sierra's fidelity by comparing current status with that of implementation. As seen on slide 30, these metrics should be measured at program implementation and again at program evaluation, and ideally more frequently.

The Coverage metric examines the percent of the target population that is served by the program currently as compared with the percentage covered at implementation. As you remember from the previous slide, Program Sierra's leadership is interested in the percent of the target population being reached, or the program's coverage.

The Content metric assesses the activities that were conducted at implementation as compared with those that are currently being conducted. Have additional activities been added to the program or have program activities been discontinued? If there have been changes to the program activities, what was the purpose of the change and how did it change?

The Frequency and Duration with which the program's activities are currently conducted are compared with the frequency and duration of the activities at implementation. Were there any changes? Were the frequency and duration of the activities in line with the content and delivery of the activity?

Questions for each metric should be reviewed and documented at program implementation and at least annually to analyze program processes. However, Program Sierra's evaluation team should update the information more frequently if new program activities are added, eliminated, or are modified during the year.

Maintaining this information will be relevant when analyzing the program outcomes as it can provide insight as to why Program Sierra may or may not have had an impact on outcomes.

Next, we'll review examples of process measures and quantitative analyses that have been conducted as part of Program Sierra's process evaluation.

[Slide 31]

On slide 31, we review Coverage and Participant Demographics metrics that were analyzed using Program Sierra's administrative database. While this database contains data on the number of program participants and their demographic characteristics, the target population is estimated using research data.

When comparing the number of individuals that participated in Program Sierra to those that were targeted, we can see that the program is only reaching 87% of the targeted population as a whole, and that only 70 percent of females are being reached by the program.

Coverage and demographic information should be collected to learn whether your program is reaching the intended audience. Your program should also be able to identify the number of participants it is reaching and whether those participants are new or returning to the program.

[Slide 32]

Process metrics for recruitment, retention and participant return can help you track new or

returning participants and ensure proper coverage and service for your program.

Slide 32 is included as an example of visually portraying a process metric examining trends in recruitment, retention, and return participants. As you can see, the number of participants recruited and retained by Program Sierra increased substantially during the first quarter (Jan-March) but then tapered off during the second quarter (April-June). However, during the same timeframes, the number of returning participants spiked in March, fell below all previous months in April and appears to have generally leveled off during the second quarter.

Tracking program recruitment, retention and/or return can also help Program Sierra identify opportunities for improvement. Interviewing participants who leave the program before completion, for example, can identify potential program improvement opportunities.

Now let's see how satisfied Program Sierra's participants are.

[Slide 33]

To analyze satisfaction, Program Sierra's evaluation team distributed a questionnaire among program participants. This questionnaire simply asked; "How satisfied are you with the services offered?" and measured responses on a five point scale where 1 = Not at all satisfied, 2 = Somewhat satisfied, 3= Neither Satisfied nor Dissatisfied, 4 = Very satisfied and 5 = Extremely satisfied. A frequency count analysis of this question can be seen on slide 33. The counts for each response option are shown in the second column entitled "Number," followed by the percentages in the final column.

Your interpretation of the percentages may be enhanced by combining response options. For example, combining the percentages for the Extremely and Very Satisfied response options tells us that 45% of respondents were very or extremely satisfied. Likewise, combining the percentages for the remaining response options indicates that 50% of the program participants were not at all, somewhat, or neither satisfied or dissatisfied with the services. Note that 5 percent of the participants did not respond to the question. Non-responders should be included in your table but should never be combined when response options are being collapsed.

With half of the participants not being satisfied, you would want to keep this in mind if your outcome analyses, which we will discuss in the next presentation, do not reveal any changes in program participant outcomes. A possible relationship between low satisfaction and poor program outcomes illustrates just how process analyses can impact outcome analyses.

One other note to mention, you may choose to eliminate the "Neither Satisfied or Dissatisfied" response option, thus requiring program participants to make a decision as to their satisfaction with the program. This is considered a forced choice option.

The next slide illustrates an examination of the frequency of Program Sierra's activities. As you may recall from slide 30, changes in the frequency and duration of program activities are metrics that can be used to evaluate the fidelity with which program activities are implemented.

[Slide 34]

Slide 34 illustrates an analysis of Program Sierra's activities by comparing their current frequency with that at implementation.

You can see that Program Sierra conducts at least four activities:

- Psychological health screening,
- Outreach,
- Resilience education, and
- Research

From when the program was implemented to now, there have been no changes in the frequency of the outreach and resilience education activities. However, changes have occurred in the psychological health screening and research activities.

Let's look at the change in psychological health screening. When Program Sierra was implemented, these screenings were conducted at every visit, but they are currently being conducted at the initial visit and every 60 days. The reason for this change is because the scientific evidence no longer recommends psychological health screenings at every visit.

This change in frequency could have an impact on participant outcomes. You can examine whether it does by comparing the outcomes of participants who received screenings at every visit with the outcomes of participants who received screenings at initial visit and every 60 days. Remember to ensure the two groups of participants being compared are similar in characteristics such as age, gender, rank, ethnicity, service branch, and participation. Doing so will enable you to conclude that any differences in outcomes are more likely due to the change in frequency of the program activities.

[Slide 35]

The examples we just reviewed illustrate how process analyses, such as participant satisfaction and knowing how the program may have evolved or changed over time, can affect participant outcomes.

Whether the program was implemented with fidelity (or not) provides insight into whether the activities implemented are affecting outcomes, which can help determine the extent to which outcomes may be attributed to the program.

Now, Mr. Frank will discuss outcome analyses.

[Slide 36]

Thank you, Dr. Best. The purpose of this section is to provide an overview of analyzing and interpreting program outcomes as part of an evaluation effort. Outcomes are essentially the end result of a program's processes. They indicate whether a program has produced changes in participant's functioning, behavior, attitudes and knowledge, consistent with the program's objectives.

[Slide 37]

The most critical question to be addressed by outcome analyses is whether a program achieved its intended outcomes, or in other words, did the program do what it set out to do? The answer to this question is critical in guiding program improvements as well as in maintaining results over time.

Other questions of interest that can also guide program improvements include: Did the outcomes vary by sub-population or perhaps according to which interventions they received? If

so, this may mean that the program's practices should be modified to better meet the needs of certain groups or that certain components within a program should be emphasized over others.

Were there any unexpected positive or negative effects of the program? If unexpected positive effects occurred, this may mean the program offers more than it anticipated. Such findings should be communicated to stakeholders who love getting more than they bargained for. If unintended negative outcomes are discovered, such as privacy violations or perhaps injuries resulting from a training activity, then the program will need to adapt its practices to better balance risks against potential benefits.

[Slide 38]

For outcomes analyses, and generally, any other type of analysis, it's best to start by getting a sense of your data. Response frequencies will allow you to examine the shape of your distribution of scores. For example, in the graph on the left on slide 38, you will notice that higher scores, such as 8, 9, and 10, occur more frequently than lower scores. This indicates that higher scores were more common among the participants, as reflected in the shape of the distribution being skewed to the right. This means that the average score will be pulled toward the upper end as well.

Group averages, as shown in the figure on the right, give you a sense of how outcome scores varied across participants of different age categories. You might expect, for instance, that younger participants would score higher on average for some measures like physical fitness; whereas, relatively older participants might score higher on other measures based on experience or knowledge.

[Slide 39]

Next, begin to address key outcome evaluation questions, which may differ somewhat depending upon the type of program and its objectives. As mentioned before, the core question is whether the program achieved its intended outcomes.

Displayed here, we show a program focused on three outcome domains: Quality of Life, Resiliency, and Job Functioning. The outcome domains of Quality of Life and Job Functioning each have one outcome measure, while the outcome domain of Resiliency has two. Analyses should be tailored to match the type of data derived from each of these measures, and in many ways this is like having four different questions in one.

[Slide 40]

Key outcome evaluation questions require a direct comparison between measured outcomes and the program's stated objectives. As we have tried to emphasize in previous episodes in this series, it is critical to begin with objectives that are SMART – specific, measurable, achievable, relevant and time-bound. If you begin with SMART objectives, it is much easier to compare measured results to what your program intends to achieve.

Even if a measured outcome does not match the stated objective, this is tremendously informative in terms of guiding improvements in the program that will ensure its sustainability. For instance, if an objective was not achieved in one year, then barriers can be identified and dealt with to enhance the likelihood of success for the following year.

The next two slides contain examples of comparisons between objectives and measured outcomes.

[Slide 41]

On slide 41, consider a clinical program with a stated objective to reduce depressive symptoms between pre- and post-treatment assessments. The measured outcome showed that, averaged across all participants, depression scores decreased from nine to five. We won't go into specific analysis strategies too much, but generally the question of whether a significant or noteworthy change has occurred is determined by comparing averages at different time-points or across different groups. That determination involves comparing the averages and the amount of error, or variability around those computed averages. In this case, the error bar for Pre-treatment scores suggests that the true average score falls somewhere between 7.4 and 10. Likewise, the error bar for the Post-treatment scores indicates that the true average score falls somewhere between 3.9 and 6. Importantly, the error bars around the averages do not overlap, so for our purposes, we can say that the averages at pre- and post-treatment assessments are indeed different and there is a clear decrease in depressive symptoms over time.

[Slide 42]

On slide 42 is an example involving a non-clinical program focused on improving resilience from baseline through a 6-month follow-up. It is a bit more complicated than the previous example because it includes three time-points. Measured outcomes revealed that average resilience ratings increased from 15 to 30 between baseline and post-program assessment but then declined to 20 by the six-month follow-up. As you can see in the graph, the error bars around the mean for the baseline score do not overlap with the error bars around the mean of the Post-program score. This allows us to conclude that increased resilience was indeed evident from baseline to post-program assessment. However, the benefit was not maintained over time. The fact that the error bars around the averages at baseline and those at six-month follow-up do overlap suggests that the scores are not significantly different.

This information may be very useful in terms of informing program improvements, which might include using additional training sessions to ensure sustained improvement or perhaps linking select participants to additional services.

[Slide 43]

Let's take a look at a different type of outcome evaluation question – whether outcomes vary across sub-populations or intervention groups. Consider, for example, a non-clinical program that focuses on increasing learning among personnel from three different service branches.

Outcomes are similar between service branch A and service branch B, but the learning score is substantially lower for service branch C. This informs program managers about the need for modifications or improvements directed toward improving learning scores for service branch C. It could be that service branch C finds the subject matter irrelevant or that they cannot fully take advantage of the program because they are unable to attend consistently.

[Slide 44]

On slide 44, consider a similar example involving a clinical program focused on decreasing posttraumatic stress disorder (PTSD) symptoms and whether there are any positive or negative

outcomes beyond the target outcome. This is very important because many outcomes of interest are interrelated – they co-occur with one another at high rates, such as posttraumatic stress disorder, depression and substance abuse.

In this case, the vertical axis on the left of the figure represents the average percent decrease in symptoms between baseline and post-treatment. The results showed there was indeed a notable decrease in PTSD symptoms – 25 percent – and also a fairly substantial decrease in depressive symptoms – 20 percent – between baseline and post-treatment assessment.

There was a smaller decrease – 7 percent – for substance abuse. This could be interpreted as an unintended benefit, given that the program wasn't heavily focused on substance abuse as an outcome. Alternatively, it could be interpreted as an opportunity for improving the program if program personnel and stakeholders think that substance abuse should be better addressed by the program. The interpretation will really depend upon the program and the needs of participants.

[Slide 45]

Reporting results to stakeholders will be the focus of a future episode in this training series but is worth mentioning here. Effective summaries of outcome analyses will include information about the following topics, closely related to the evaluation questions posed at the start of this section.

Stakeholders will definitely want to know whether intended outcomes were achieved and how program administrators know they were achieved. It is no longer enough to say, "We think we are benefiting participants." Rather, stakeholders want to hear, "The program is producing benefits, and we know we are achieving those benefits based on evidence from our ongoing program evaluation processes."

Likewise, stakeholders will want to know about multiple outcomes, including target outcomes and those that are the focus of their respective interests, such as maintaining force readiness or the ability of service members to perform their job functions.

As mentioned in this and previous presentations, measurement and analysis provide great opportunities to highlight program strengths and to develop targeted program improvements based on actual program data.

Finally, stakeholders might wish to see details about how program improvements are to be carried out. Of note, there really is a distinction between program "weaknesses" and so-called "opportunities for improvement." Simply pointing to weaknesses in a program really doesn't do anyone any good, but if a program manager can say, "We are not fully achieving our intended outcomes, but here is a list of improvements we intend to make to enhance program effectiveness," then the evaluation process has resulted in true opportunities to better serve program participants.

[Slide 46]

Analyzing and interpreting program data is a complex process, and programs vary greatly in their capacities to carry out these activities. As such, it is important to be aware of some of the most common challenges that arise, how they can potentially be addressed and what resources are available for support.

[Slide 47]

On slide 47 are questions reflecting a few of the more common concerns that have been expressed in our interactions and trainings with program managers and service leadership. The answers to these questions follow on slides 48 through 50.

[Slide 48]

How can I assess program fidelity when I have limited information of the program form when it was implemented?

This question is especially common for older programs with personnel who may have retired and records that are likely to have been archived, lost or even destroyed. There are two basic responses to this concern.

First, you may be able to locate some information in historical records or archives, interview former program personnel or check for information in service-level databases. This solution is basically to use the information that is available and do the best you can to determine whether fidelity has been achieved by comparing present processes to those at program initiation.

Second, if very limited information is available or especially if the program has changed a great deal since its inception, it may be necessary to re-initiate, or reconstitute a program in its current state with the most recent mission, goals and objectives. The re-initiated program will then serve as the baseline for future examinations of fidelity.

Importantly, if program personnel choose to re-initiate the program, this provides an excellent opportunity to revisit the evidence basis for the program using the most up-to-date research on relevant program practices. Likewise, it will be important to revisit the need or impetus for the program to ensure that the reconstituted program effectively serves its target population as it currently exists.

[Slide 49]

“How can I conduct process and outcome analyses with limited resources?”

As we have said throughout this presentation, you cannot know if your program is effective unless you collect and analyze the right kind of data. In this age of accountability, your stakeholders will appreciate your ability to demonstrate where your program is effective, and also what it plans to do in other areas to become more effective.

We recognize that many programs do not have the resources or personnel to conduct more advanced analytics such as inferential statistics. However, there are many descriptive techniques that can be performed in Microsoft® Excel software that can help you answer important evaluation questions. Frequency counts, percentages, means, and standard deviations can all be performed in Excel.

If possible, include data collection and analysis into your standard operations rather than trying to design a comprehensive study. A lot can be accomplished with a little data and the right people! Start with a simple evaluation question and build on that experience. Once you get your feet wet with process and outcome analyses, the sky is the limit!

[Slide 50]

Another common question is, how do I conduct analyses for a program that has many separate but interrelated components?

Many programs have several core components, and in general, it is best to measure and analyze program data with as much detail and accuracy as possible.

Specific measurement and analysis processes will depend upon the goals of evaluation. Given the program's logic model, it may be best to examine whether each component or major program process can be linked with its intended outcomes. This will help to determine the usefulness and benefits of each component.

Similarly, it is a very good idea to examine whether specific components are implemented with fidelity. This type of examination will help program managers to determine very specifically where opportunities for improvement exist. For example, potential improvements might include enhancements to improve fidelity or elimination of those components that cannot be realistically achieved given available resources.

Finally, general analyses of the program as a whole remain useful in terms of conveying the value of the program to stakeholders and maintaining accountability.

[Slide 51]

Thank you, Dr. Best, Mr. Frank, and Ms. Aguirre.

You've heard a great deal today about analyzing and interpreting data on program processes and outcomes which are essential to building a culture of effectiveness in the Defense Department's system of prevention and care for psychological health and traumatic brain injuries.

[Slide 52]

A key takeaway is that programs can use data analysis to firmly establish evidence of a program's effectiveness, which is critical to a program's survival and to ensuring that service members receive the best possible prevention and care. Without measurement and analysis of program data, it is not possible for programs to state with assurance that they are fulfilling their missions.

Measurement and analysis also establish linkages between a program's resources and processes and its outcomes. This ensures that program managers can demonstrate that their resources are used effectively and efficiently and that the activities and products of those resources are worthwhile in achieving program objectives.

Finally, measurement and analysis can be used to guide program improvement efforts. By examining program data, it is possible to identify opportunities for program personnel to enhance program effectiveness and to better support service members in carrying out their important missions.

I hope you will continue to attend these training presentations and also consult the Program

Evaluation Guide and other resource materials on the DCoE website.

Now, back to Ms. Aguirre.

[Slides 53-56]

Thank you Captain Thoumaian. There is a great deal of useful information available to programs on measurement tools such as surveys and checklists, and on data privacy and storage considerations. On slides 54 through 56, we provide a list of relevant references and resources that we think may be useful.

[END]