



Analyzing and Interpreting Data on Processes, Outcomes and Costs

Presented on June 17, 2014

Episode 4 in the Program Evaluation and Improvement Training Series

Presenters

CAPT Armen Thoumaian, Ph.D.
Health Science Officer
Office of Policy, Programs and Integration
Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury (DCoE)
Silver Spring, Md.

Aaron Sawyer, Ph.D.
Research Scientist
Contract support for DCoE
Arlington, Va.

Patrick High, Dr.P.H.
Epidemiologist
Contract support for DCoE
Arlington, Va.

Camille Hinds, Ph.D.
Applied Economist
Contract support for DCoE
Arlington, Va.

Moderator

Debra Stark, M.B.A.
Research Scientist
Contract support for DCoE
Arlington, Va.

[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 another episode in the 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 Centers: the Defense and Veterans Brain Injury Center (or DVBIC), the Deployment Health Clinical Center (or DHCC), and the National Center for Telehealth and Technology (or T2). DCoE and its Centers work closely with one another to promote high-quality prevention and care across the Defense Department. Together, we produce a variety of trainings on subjects ranging from program evaluation to clinical care and prevention practices.

The DCoE Program Evaluation and Improvement training series is designed to increase the capacity of psychological health and traumatic brain injury programs to engage in program evaluation activities.

The trainings in this series are directed toward program administrators and service leadership who are currently involved with or plan to conduct program evaluation activities.

This series contributes to DCoE's larger mission to enhance the quality and effectiveness of psychological health and traumatic brain injury programs by providing training on key activities that may be used to advance program evaluation and improvement efforts.

On behalf of the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury, thank you for participating in this training series.

[Slide 1]

Ms. Stark: Hello. My name is Debra Stark. I provide contract support to the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury. I will be your moderator for this presentation, Episode 4 in the program evaluation and improvement 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 and Interpreting Data on Processes, Outcomes and Costs." 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 located on the left of your screen. You can do this at any time, 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 as the connection catches up with the audio at times. Depending on your network security settings, there may also be some noticeable buffering delays.

Closed captioning is not available for this event.

[Slide 3]

Continuing education credit is not available for this event but may be available for future webinars. Webinar materials from this series will be made available in the Program Evaluation section of the DCoE website. For information about other DCoE webinars and trainings, visit the Training section of the DCoE website by following the link on slide 3. Slides and other materials are available in the boxes at the bottom of the screen during the webinar.

[Slide 4]

This webinar was introduced by Captain Armen Thoumaian. Captain Thoumaian is a Health Science Officer with 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 completed a National Institute of Mental Health fellowship in Community Mental Health.

[Slide 5]

Presenters for this episode include Dr. Aaron Sawyer, a Research Scientist providing contract support to DCoE. Dr. Sawyer is a clinical psychologist with extensive expertise in intervention outcome research and program evaluation. He has delivered child, family, and adult interventions for more than a decade, including specialization in trauma and experience working with military families. Dr. Sawyer holds an M.S. in Experimental Psychology and a Ph.D. in Clinical Psychology. He completed postdoctoral training at The Kennedy Krieger Institute of Johns Hopkins University and is a Licensed Psychologist.

Our next presenter is Dr. Patrick High, an epidemiologist providing contract support to DCoE. He has over a decade of experience and expertise in survey design, research methodology and program evaluation. His experience includes supporting the Office of the Undersecretary of Defense for Personnel and Readiness, Operations Research and Safety, and the Defense Suicide Prevention Office as an epidemiologist. Dr. High holds the degree of doctor of public health with specialization in Epidemiology and Biostatistics from the Uniformed Services University of the Health Sciences and previously spent nine years in the Illinois Army National Guard.

[Slide 6]

Our third presenter is Dr. Camille Hinds, an applied economist providing contract support for DCoE. Dr. Hinds holds a doctorate in economics from American University specializing in advanced statistical analysis, econometric modeling and economic evaluation. Her technical expertise has been applied to home mortgage defaults, obesity outcomes, food program participation and military programs. She is an expert in analyzing complex data sets and large national surveys as well as developing survey instruments. She has also taught numerous courses in economics and statistics and has over seven years consulting and government contracting experience.

Dr. Sawyer: Our final presenter is Ms. Debra Stark, also a Research Scientist providing contract support to DCoE. Today, Ms. Stark is both moderator and presenter. She is a survey methodologist and analyst with more than 15 years of research experience. Ms. Stark's work includes program evaluation and monitoring, qualitative data analysis, and survey instrument design. She has worked on health services evaluation projects with several Federal agencies, including the Department of Veterans Affairs and Tricare Management Activity. Ms. Stark received her MBA from Vanderbilt University.

[Slide 7]

Ms. Stark: This training presentation will provide guidance on analyzing and interpreting program data on processes and outcomes for program managers and others involved with program evaluations. It will also provide an overview of strategies used to analyze program costs. Topics will include an introduction to general strategies and concepts integral to analyzing and interpreting data on processes and outcomes, fundamentals of economic analyses, and consideration of common challenges that arise when analyzing and interpreting data.

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

- Explain general strategies and key concepts relevant to analyzing program data on processes and outcomes
- Demonstrate basic knowledge of economic analysis strategies relevant to program data
- Implement suggested guidance to initiate analysis and interpretation of program data
- Identify common challenges that programs face in analyzing and interpreting data and resources for technical support

[Slide 8]

As seen on slide 8, the first topic is an introduction, followed by an examination of process analyses, outcome analyses, and a unit on analyzing program costs. Finally, we present common challenges, concluding comments and resources, and we will end with an opportunity to provide Feedback and a live Q&A session.

I will begin the presentation with a very brief introduction, followed by Dr. High, Dr. Sawyer and Dr. Hinds, with concluding comments from Captain Thoumaian.

[Slide 9]

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 10]

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 resources, references and reading recommendations at the end of the presentation, on slides 66 through 68, for those who want to learn more.

[Slide 11]

Analysis enables us to describe large amounts of information and discover patterns in the data.

In order to increase the knowledge base for psychological health and traumatic brain injury and to promote high standards of prevention and care, we need to focus on interventions and practices that have the greatest evidence of effectiveness. We need to understand whether a program is working. This involves collecting information, analyzing data, and interpreting what the data mean.

[Slide 12]

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 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 to creep in. Coding mistakes are a major source of error in survey measurement. For example, an answer of 5 on one question means that it was rated "the best," or we want more of that thing, while an answer of 5 to another question means it was "the worst," and we'd like less of it. The data will need to be re-coded so that all answers are in the same direction and can be summarized and considered together.

When responses to an item are so few as to make the analyzing and reporting on them meaningless, they will need to be combined with other categories, and the category title or label shifted accordingly. Generally, there should be 5 or more responses to a question to avoid what is known as "the small cell problem," and so that a participant could not possibly be identified by someone combing through the data.

Look at summary responses to the items on your measures to be sure they make sense: if you have results for 400 program participants, the answers to "What is your gender" should not total to 500. Are certain answers about what you expect? Do some answers seem too high or too low?

[Slide 13]

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

Data analysis can help you learn what is typical for your program, and can help you find the spread or variation among participants, processes, cost and outcomes.

Generally, statistics are classified into two types: descriptive and inferential. Descriptive statistics summarize and can show relationships between variables. Inferential statistics enable one to generalize or infer findings from one sample to a larger population. We will only touch upon those.

[Slide 14]

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.

In some cases, a numerical count is all you need or want. In other cases, counts serve as a base for other calculations, such as for percentages, which express data as a proportion of a whole. Even when working with percentages, it is common to include the count: the total number in a population (uppercase or big “N”) or a sample or subgroup from within the larger population (lower case, or little “n”).

[Slide 15]

Reporting percentages is a very useful way to think about numerical data: we are all familiar with percentages and they are readily understood. A few rules are important to follow when working with percentages.

First, use the correct denominator. The denominator is the base from which the percentage is calculated. Use the right base, and communicate to others which base you are talking about. 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? This is very important when a large number of responses are “missing” or marked as “not applicable”: if you use the total number of participants as a base, the percent may not be correct. When reporting percentages, such as answers to a survey, show how many “missing” items there were.

Second, even if a software program such as Microsoft Excel calculates it for you, you do not need to report results out to the 4th decimal place. In fact, it is a bit misleading to do so, since it implies a more precise result than what you may actually have, such as 6 people saying ‘yes’ and 17 saying ‘no.’ In general, use just one or two decimal places, and round up any number 5 or greater.

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

Lastly, when you do add up percentages, be careful that you do not take the next step and sum up several percent responses and divide to obtain a percent average: errors can creep in. Always go back to the original numbers. Often the errors are small, but at other times they can be larger and lead to greater misinterpretations.

[Slide 16]

Measures of central tendency tell you what is typical for a distribution of scores or for a group. The most commonly-used of these measures are the mean, median and mode. The definitions are presented here. They let you know the “typical” value for a group. 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. Mode is the most commonly occurring value.

[Slide 17]

In addition to measuring the central tendency of a distribution, it is also common to report the variability: the dispersion, or spread of a distribution. The range expresses the distance between the lowest and highest scores. On slide 17, these two datasets have the same mean, but you can see that participant scores for Program 2 varied more widely than those for Program 1. The range gives information about extreme scores. It ignores other information about distribution, which is represented in the standard deviation.

[Slide 18]

The standard deviation provides information about how far away from the mean the data fall, on average. It makes use of the information about every score, as we see represented in the center of slide 18, 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, while on the upper right, we see that most responses varied greatly from the mean. When all answers 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, variation may indicate a less-than successful outcome.

[Slide 19]

For your analysis, you will want to inspect the data using the techniques discussed so far: numerical counts, percents, mean, median and mode, as well as the variability measures of range and standard deviation. We encourage you to examine the data and relate what you find to evaluation questions, check for data quality or errors and to see whether the assumptions underlying analysis are met, such as a normal or Bell curve distribution. These determinations will guide choices regarding inferential analyses that will allow you to evaluate the size and scope of effects and whether your results generalize beyond the immediate context of your evaluation.

[Slide 20]

Check the data for patterns. The answers to some questions may seem to link with responses to other questions. Work with the data and delve into this further. It may be helpful to re-frame your data into pie charts, bar charts, tables, lists and line graphs. See how the data look when

displayed differently. Does anything 'jump' out at you? Using charts and graphs may help you organize your report-writing, and using them may help you to communicate most effectively with others.

Crosstabs, a form of sorting, will allow you to get an overall picture of what is going on. For example, do satisfaction ratings for a program vary by age, rank or education? Crosstabs can readily be created using two categories, allowing you to examine subgroups at an even deeper level. When you find patterns, you want to investigate them further using inferential statistics.

[Slide 21]

Just because something looks like an important finding does not make it so. Inferential statistics enable one to generalize or infer findings from one group to a larger population and see whether that relationship is truly meaningful. Relationships found in one group will not necessarily hold true for the wider population. Measures may be applied to see whether a relationship is negative or positive, the strength of that relationship, the likelihood of that relationship occurring again, or whether some finding was simply due to chance. This type of analysis will require some expertise, possibly involving outside resources.

[Slide 22]

In the world outside of academic research, we often do not have the luxury of a control group to establish whether an intervention specifically caused changes in participants. However, in some cases comparison groups may be available for program evaluations. Again, some statistical expertise will be required to establish that the groups vary only in whether they receive the 'intervention,' that is, the program's clinical, education, outreach, or ancillary component.

[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, for the next portion of the presentation, I will turn to Dr. High.

[Slide 24]

Dr. High: Thank you Ms. Stark for the introduction and providing an overview of data analysis. In this section of the presentation I will review process analyses.

[Slide 25]

As seen on slide 25, when conducting process analyses, there are several questions you will want to consider. Primary among these questions is: was the program implemented with fidelity? This is an important question to answer as programs generally are based on scientific evidence and have a structured process intended to create changes among program participants. Other questions of interest may include: How does the program operate? What is the program expected to achieve?, How is the program expected to achieve what it has set out to achieve?, and finally How did 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?” Program processes are directly linked to program outcomes and facilitate the measurement of these outcomes, which will be discussed in the next section of the presentation.

[Slide 26]

The information contained for each of the questions seen on the previous slide can generally be answered by having an understanding of the components of process analysis and the program. These components include:

- 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 of the program and how did those barriers affect the program? Barriers can include inadequate funding at implementation, reduced funding during program operation, and lack of training of program staff, or not having the number or type of staff needed to implement and run the program.
- Services/activities – what activities are being conducted by the program and are they in line with the program’s mission statement, goals and objectives, and aligned with the program’s logic model? Services and activities can include clinical, outreach, education, and/or research/ancillary activities that are to have some effect on program participants or the wider group of individuals that may be impacted by the information being provided.
- Exposure – who has access to and/or is exposed to the program, how are these individuals exposed to program, and does the program have a recruitment or retention strategy to ensure it continues to receive participants or follow participants to ensure program participation as prescribed?
- The final component listed, Context, includes the environment in which the program is actively taking place and how environmental factors, such as service mandates or directives, or change in mission may impact the program and thus cause the program to change over time.

While these components are for consideration, we will want to analyze our core question.

[Slide 27]

The primary question when conducting a program analysis is: Was the program implemented with fidelity? While other questions are important, the answer to this specific question will likely have a bearing on whether or not the program has been shown to change program participant outcomes. Slide 27 includes a matrix for how to analyze program fidelity.

The fidelity metrics to consider include; Coverage, Content, Frequency and Duration. Each of these metrics should be measured at program implementation and again at program evaluation, and ideally more frequently.

For coverage, the percent of the target population that was covered by the program should be compared to the percent that is being covered by the program currently. Has the proportion of the population being covered increased or decreased. Why have these changes occurred, if any?

The content includes the activities that were being conducted at implementation and the activities that are currently being conducted. Have additional activities been added to the program or have program activities been stopped? If there have been changes to the program activities, what was the purpose of the change and how did it change?

Next, you will want to analyze the frequency and duration of each activity from program implementation to currently. How frequently were program activities being conducted and for how long were these activities being conducted? Were there any changes, and was the frequency and duration of the activities in line with the content and delivery of the activity?

Each question for each metric should be reviewed and documented at program implementation and at least annually to ensure program processes can be analyzed accordingly. However, you should update the information more frequently if new program activities are added, cease to exist, or are modified during the year.

Maintaining this information will be relevant when reviewing the program outcomes and provide insight on why the program may or may not have had an impact on outcomes when the outcome analysis is conducted.

Next, we'll review examples of process measures and quantitative analysis that has been conducted.

[Slide 28]

As seen on slide 28, there are three processes with several examples of metrics that fall within each one. The metrics displayed in red will be used in the examples in the slides that follow.

The first process to track, participation, refers to calls to a helpline, session attendance, number of sessions held, the targeted population, participant demographics (for example; rank, branch of service, sex) and perhaps how the participant was referred to the program.

The second process to track is program satisfaction, which includes satisfaction ratings from evaluation cards or other evaluation processes provided to program participants, such as ratings of how likely a person is to refer other individuals to the program.

Finally, activities include the frequency and length of each program activity conducted, the number and type of activity and should also include the number of referrals made.

[Slide 29]

On slide 29, we'll review Coverage and Participant Demographics which is related to our earlier slide regarding Program Fidelity.

When comparing the number of individuals that participated in the program to those that were targeted, the program is only reaching 87% of the targeted population. Similarly, when reviewing

the branch of service, coverage ranges from 85 to 93 percent, but only 70 percent of females are being reached by the program.

This level of coverage may or may not be sufficient for the program, but without the analysis and an understanding of the programs mission and goals, there would be no way of knowing. Rarely does a program reach an entire population, but the program should be able to identify the number of participants it is reaching and whether those participants are new or returning to the program.

Coverage and demographic information should be collected to learn whether the program is reaching the intended audience; the program should also collect satisfaction information from participants as appropriate.

[Slide 30]

A basic level of satisfaction information should be collected from program participants. Such information can ask, "How satisfied are you with the services offered?" Satisfaction can be measured 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. When such a scale is used, a count of each response should be provided as can be seen in the second column titled Number. The percent responded for each category is then provided.

You may notice that when you review the proportion column, each category response may not provide you with a compelling story of participant satisfaction. Because of this, you should combine response options. In this example, Extremely and Very Satisfied were collapsed into one category while the remaining response options are collapsed into a second response option thus making two response options.

When response options are combined, you can see that 45% of respondents were extremely/very satisfied while an almost equal number (50%) were not as satisfied. These two categories account for only 95 percent of the participants, while 5% did not respond to the question. Non-responders should be included in such a table but should never be combined when response options are being collapsed.

You should be aware of the proportion that does not respond. In this example the low number of non-responders does not present a concern for this example.

With nearly half the participants being satisfied and half not being satisfied, you would want to keep this in mind if the program did not see changes in program participant outcomes. While lack of satisfaction and a lack of positive outcomes do not indicate a true cause and effect relationship, there could be a relationship between low satisfaction and poor program outcomes.

Additionally, you may choose to modify the response options and eliminate the "Neither Satisfied nor Dissatisfied" option, thus requiring program participants to make a decision as to their satisfaction with the program. This is considered a forced choice option.

[Slide 31]

In this non-clinical example, "How did participants perceive the usefulness of the training session?" you will notice that the response options required participants to choose a response as there was no 'neutral' option for consideration. Similar to the previous example, the four

response options were combined into two; Extremely/Very Likely and Somewhat/Not at all Likely. Where 1/3 of program participants felt the training session was useful while, nearly 2/3 (63%) did not perceive the training session to be useful.

Again, being aware of this information will be important when reviewing program outcomes.

Once the program coverage and participant satisfaction is known, you would want to review the frequency and duration of program activities, which I will discuss in the next two slides

[Slide 32]

On slide 32, the frequency of program activities are analyzed. You can see that the program conducts at least four activities: psychological health screening, an outreach activity, resilience education and research.

From when the program was implemented to now, there have been no changes to how frequently the outreach activity and resilience education is provided to program participants. However, changes have occurred in the psychological health screening and research activities.

When the program was implemented, psychological health 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 for conducting the screening has changed and does not need to occur at every visit. This modification may or may not have an impact on program and participant outcomes, but the only way to know is to compare program participant outcomes of those that are receiving the updated screenings to those who received the screenings when implemented.

When analyzing the research activity, it appears the program did not conduct any research at the time of program implementation, but it is currently comparing participant outcomes to baseline scores to determine program effectiveness.

Determining the effectiveness of the program should have been included at implementation, but it is good to see the program is currently conducting such analyses. While this program may have implemented the program as intended, it has evolved since implementation and this evolution needs to be taken into consideration when attributing changes in participant outcomes to the program.

[Slide 33]

In this non-clinical program example on Slide 33, the program is only conducting three activities; Resilience screening, outreach and resilience education.

The duration of the resilience screening has been reduced from the standard 15-minute questionnaire to a 5-minute questionnaire. Similarly, the resilience education that was previously provided via a one hour instructor-led presentation has been reduced to 45 minutes via a self-service webinar. Both changes have been made because there is fewer staff available to conduct the screening and education.

Because the activity modifications have been documented during the course of the program, these modifications should be reviewed when program and participant outcomes are being analyzed. Changes to program activities may result in diminished outcome size or reduced

duration program effects, because they did not receive the services as originally intended by the program.

As you can see, it is important to understand how the program may have evolved over time and knowing this information will be imperative when determining whether or not the program has had an impact on participant outcomes.

[Slide 34]

For the program, it's a good idea to track recruitment, retention and participant return monthly, quarterly and/or annually. Tracking such information will facilitate a better understanding of program participation.

As can be seen on slide 34, the visual representation of the data is more compelling than if it were presented in a standard table. As you can see, it appears the number of participants recruited and retained by the program increased substantially during the first quarter (Jan-March) and is slowly rising 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.

Such fluctuations may be appropriate for a program, but you will not know if you do not continuously collect, graph and monitor program recruitment, retention and/or return. Collecting this data will allow comparisons to be made year-to-year and provide the opportunity to determine if any seasonal variations are present.

As process information is collected and analyzed, the link to program outcomes and whether the program is meeting its objectives should always be considered.

[Slide 35]

The purpose of this section was to provide an overview of process analyses as they help determine the extent to which outcomes may be attributed to the program.

Whether the program was implemented with fidelity (or not) provides insight into whether the activities implemented are affecting outcomes, which can help prevent false conclusions about program effectiveness.

Now, Dr. Sawyer will discuss outcome analyses.

[Slide 36]

Dr. Sawyer: Thank you, Dr. High. 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.

However, there are other types of more nuanced questions that can also guide program improvements. Did the outcomes vary by sub-population or perhaps according to which intervention components 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 results are generalizing beyond the program's most specific objectives, and such benefits can be emphasized to stakeholders, who love getting more than they bargained for. If unintended negative outcomes are occurring, 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. For example, in the graph on the left on slide 38, you will notice that there are more responses on the upper end of the distribution, meaning that higher responses were more common and the mean 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 outcomes varied across different types of participants. 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, and each program will focus on different sets of outcomes.

Displayed here, we show a program focused on three domains with four measured outcomes: one each for quality of life and job functioning and two measures of resiliency. 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]

Each of these key questions involves directly comparing measured outcomes to stated objectives. As we have said in each episode 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 a program says it will achieve.

It is likely that measured outcomes will not match stated objectives in every case at every time point. 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 to success can be identified and mitigated 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 averages. In this case, the error bars around the averages do not overlap, so for our purposes, we'll 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, this is a slightly more complicated example involving three time-points and a non-clinical program focused on improving resilience from baseline through 6-month follow-up. The actual measured outcome was that average resilience ratings increased from 15 to 30 between baseline and post-program assessment but then declined to 20 by six-month follow-up. Because the error bars do not overlap, we might say increased resilience was indeed evident from baseline to post-program assessment. However, the benefit was not maintained over time, since the averages at baseline and follow-up are not substantially 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 group. 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 outcomes 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 due to barriers related to consistent attendance or technology issues.

[Slide 44]

On slide 44, consider a similar example involving a clinical program focused on decreasing 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 own roles, 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 are a list of improvements we intend to make to improve the program and achieve them in the future," then the evaluation process has resulted in true opportunities to better serve program participants.

Now, I'll hand off the presentation to Dr. Hinds, who will discuss analyses of program costs.

[Slide 46]

Dr. Hinds: Thank you, Dr. Sawyer. Today I will discuss some fundamental aspects of analyzing a program's costs. My discussion will reference the benefits to conducting cost analysis, define some frequently used terms, provide an overview of the most popular types of cost analyses and present the basics to conducting an analysis.

Let's begin with core questions.

[Slide 47]

Analyzing program costs is a necessary step towards evaluating the efficiency and effectiveness of a program's operations. For the analyst gathering information to support this analysis, we recommend considering the following core questions:

- For the stakeholder, the core question to be addressed by this analysis is: Which is the most effective intervention to fund?
- For program administrators, the core question to be addressed by cost analysis is: How are program funds being spent?

Other questions of interest are:

- What types of analyses use cost measures?
- How are costs quantified?
- What are the average cost values per participant?

At the root of each of these questions is the fundamental query: How can resources be used more effectively?

[Slide 48]

Before tackling how to analyze program costs, it is useful to consider the benefits to analyzing program costs.

Cost evaluations help to assess the gains and the costs to carrying out program operations. This is important because resources are scarce and stakeholders have to choose among viable alternatives. When we analyze program costs we provide stakeholders with objective measures to help make informed funding decisions.

Moreover, an objective cost analysis helps when making comparisons because often the best program choice to fund is not obvious when programs differ in terms of their services delivered, population addressed, and/or outcome metrics used.

Cost evaluations also help program administrators to track how budgets are allocated across activities, and how well the program is functioning relative to its target goals and operating budget.

In short, cost analysis can help answer questions such as: Which program provides more "bang for the buck"? How are costs allocated across program activities? What level of additional resources may be required to fund expanding the program?

[Slide 49]

The term cost can have different meanings depending on the context in which it is used. When conducting cost evaluations cost refers to the value of resources used to deliver services.

By the "value of resources used," we mean that cost measures should capture the value of all goods and services used to conduct program operations. In particular, cost measures should include both the actual amount of funds spent on resources and the value of resources used but not paid for by the program.

For cost analysts, when gathering cost information, it is helpful to think of program operations in terms of the general resource categories needed to deliver services. For example, consider that a program may define the relevant resource categories to include:

- Noncontract labor, which includes the total salaries, wages, and benefits paid to employees for time spent performing program activities. Examples include administrative staff and program personnel. Timesheets will help quantify the amount and value of an employee's time spent on multiple activities within the program.
- Contracted services, which include the costs for program activities provided by entities outside the program such as external consultants, data warehousing or physician services.
- Building and facilities costs, which include rental payments or mortgage payments, building maintenance, and operating costs such as utilities, taxes, insurance and cleaning staff.
- Materials and supplies, which includes the costs for equipment to support program activities like computers, phones and printers, and
- Donated resources, or resources not spent out of the program's budget, but which are still a part of the program's health intervention costs. These resources are valued at the amount of dollars the program would have needed to spend had they not been available for free use. Two common examples include donated facilities and volunteer labor. In the case of a program using a building free of charge, the value of this building could be estimated from current real estate values on monthly rents in the surrounding area. In the case of volunteer labor, the value of volunteer labor could be estimated from market salary or average salary figures

In an effort to gather valid and useful cost information, we recommend that program cost information be captured through bills, receipts, contracts, wages paid and value estimates.

When collecting information, for contracted resources, the analyst should quantify costs using actual expenditures rather than budgeted line items because budgets may not reflect the amount of monies actually spent.

For donated resources, the value of resources donated but not paid for out of the program's budget should be included in cost measures. Costs can be quantified as the expenditure the program would have incurred to secure the resource had it not been donated.

[Slide 50]

The three main types of cost-analyses being conducted are cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis.

Cost-effectiveness analysis is a type of economic evaluation that examines the costs and outcomes of alternative intervention strategies. In this method, the analyst evaluates the tradeoff between the health benefits and costs of one intervention relative to another intervention.

It is important to note that to enable a comparison, the outcomes from both programs must be captured using the same scale and metrics. Cost-effectiveness analysis summarizes the value of a program into a single measure that reports cost per unit of health benefit.

In other words, cost-effectiveness analysis quantifies a program's costs in dollars and quantifies a program's outcomes in nonmonetary units such as "reduced instances of suicidal ideation,"

“increased days of exercise” or “life years gained.”

In cost-effectiveness analysis a ratio is computed that reports the difference between the cost of the intervention and the cost of the alternative in the numerator and the difference between the health outcome of the intervention and the health outcome of the alternative in the denominator. This ratio answers the question: what is the extra cost to get the extra effectiveness?

[Slide 51]

The second type of cost analysis is cost-utility analysis. Cost-utility analysis is a special case of cost-effectiveness analysis in which program benefits are expressed in terms of a specific outcome measure called quality-adjusted life years, also known as QALYs. QALYs represent a measure of the value a person places on a life lived in good health.

Quality-Adjusted Life Years provide a standard scale by which to compare outcome measures which differ across various interventions. Thus, whereas cost-effectiveness analysis requires programs to use the same metric and scale to enable comparisons, cost-utility analysis allows the analyst to compare the effectiveness of various interventions across diverse diseases and conditions when outcome measures differ.

The cost-utility ratio replaces the difference in health outcomes in the cost-effectiveness ratio with the difference between the quality-adjusted life year of the intervention and the quality-adjusted life year of the alternative.

An important limitation regarding cost-utility analysis is that: In the US, the Patient Protection and Affordable Care Act, also known as ACA, forbids using QALYs “as a threshold to establish what type of health care is cost effective or recommended.” The ban on using cost-per-QALY thresholds may reflect long-standing concerns that the approach discriminates on the basis of age and disability. The worry is that the metric unfairly favors younger and healthier populations that have more potential Quality-adjusted life years to gain.

[Slide 52]

The third type of cost analysis is cost-benefit analysis. Cost-benefit analysis is a technique that expresses the program’s costs and benefits entirely in dollar terms.

Cost-benefit analysis is more complicated to apply to mental health care than cost-effectiveness analysis because it requires attaching dollar values to outcomes that are not directly measured in dollars, for example, sense of community or depression.

It has been recommended to avoid using cost-benefit analysis to address health care questions because it involves placing dollar values on life which is controversial in health discussions. Placing dollar values on an individual’s life could bias effectiveness conclusions away from programs benefitting the elderly or the young and towards individuals who make more money and thus have higher market productivity values.

[Slide 53]

The three types of analyses are summarized for your convenience on slide 53. As can be seen from the table, these analysis types differ only in how they consider the outcome measure.

For cost-effectiveness analysis, outcome measures are considered in discrete, non-monetary units such as life-years gained or reduced days depressed. For cost-utility analysis, the outcome metric used is Quality-Adjusted Life Years or QALYs. For cost-benefit analysis, only the dollar-value of the outcome measure is used.

[Slide 54]

Collecting good cost data is at the heart of every useful cost analysis. For program administrators considering collecting their own cost data, we recommend instituting a standard repeatable process to gather cost information. We suggest the following steps to quantify program cost information:

1. First, list the main activities the program performs. Note that after completing the program logic model, as described in a previous webinar, the program administrator will have a clear sense of the activities and services the program delivers, therefore this information will be easily available.
2. Second, list the resource categories used to support each program activity. Specifically, you should review each resource category: labor, contracted services, building and facilities, materials and supplies and donated resources, and determine whether it supports that particular program activity.
3. Third, assess the data available from existing sources such as times sheets, payroll accounts, bills and contracts.
4. Fourth, collect and document the activity costs on a worksheet. We recommend creating a standard template to record program resources used by program services delivered. We provide an example on the next slide.
5. Fifth, compute average cost values. Once program cost information is recorded on the cost worksheet, it is straightforward to compute average cost values by activity and overall.

Knowing the cost of each activity helps answer the questions about the total cost of the program and the costs of specific activities.

[Slide 55]

We recommend developing a cost worksheet to help organize and standardize the process of recording program cost information. We provide an example of a worksheet.

A useful cost worksheet should include:

- a title: such as “Total Program Costs”
- the Program name “Military Resiliency Training”
- the Time period that costs represent
- the Cost Components and Resources Categories
- the Program Activities and Services provided, and
- total values

In the best case scenario, costs should be listed by the resource activity and program service. For example note that the worksheet illustrates how the total expenditures on materials and supplies were allocated across the four key program activities. In this way, the cost analyst will be able to compute summary values for each activity and each resource category.

For this sample program, we see that:

- this program spent over \$169,000 on labor costs where almost half of this amount came

- from administrative costs.
- that the largest value of donated resources was allocated to educational activities, and
- the most expensive resources were labor and facilities costs and the most expensive activities were Education and Administrative.

[Slide 56]

After collecting and recording program cost information, analysts can compute average values to address per unit questions. Here again we have a sample worksheet to record average program costs. On this sheet we have added the number of program participants.

Cost per participant is recorded as the ratio of the total cost values from the cost worksheet divided by the participant count. In this example, we see that the cost to this program for delivering education services to one thousand participants is \$132.90 per participant.

It is clear that once cost values have been recorded, it is a straightforward exercise to compute average values by participant for each key activity and overall.

Now, we return to Dr. Sawyer.

[Slide 57]

Dr. Sawyer: Thank you, Dr. Hinds. 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 most common challenges that arise, how they can potentially be addressed and what resources are available for support.

[Slide 58]

Within the military's psychological health and traumatic brain injury programs, common challenges include difficulties obtaining records, which are especially common for older programs that may have archived their materials.

In addition, programs change over time, and a program by a given name in the present may be fundamentally different from when it began. This can make any comparisons of program data over time very challenging.

It can also be challenging to capture all relevant data and to avoid capturing too much data, which can burden individuals responsible for collecting and analyzing data. As such, program managers should make sure they know which processes, outcomes and costs are most important to assess, including those that interest stakeholders.

An additional challenge relevant to younger programs is that impact, or the degree to which a program affects a population, cannot be determined immediately. Evaluations of impact often take several years and coordinated efforts between programs and those who collect population-level data.

Finally, some costs may be difficult to determine for military programs, because they are funded from multiple sources, or because complex procedures are required to obtain information about costs.

[Slide 59]

On slide 59 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 60 through 63.

[Slide 60]

How can I assess program fidelity when I have limited information from program initiation?

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 61]

“What are some effective ways to address attrition with regard to validity of a program’s outcome data?”

This is a very complex question and depends on when the attrition was first noted. If participation has concluded and you have completed data collection, you may be able to account for the attrition during the data analysis process as there are some standard methods that can be applied.

If you are collecting information and you are checking your data as you go and notice patterns of information loss, then other measures may need to be taken, up to and including stopping the program or a measurement procedure if the loss seems too severe.

Taking a very proactive approach before and during an evaluation is always the best way to manage attrition and related concerns. As such, program managers and evaluators should focus on:

- When attrition is occurring – In some cases, attrition is due to non-completion of certain activities, while in others participants may be unavailable only during follow-up data collection procedures that occur after program completion.

- It is also important to determine the cause of attrition, if known. For example, were program participants lost to follow-up or left the program because they died, because they have moved or perhaps for some other unknown reason?
- Finally, it is important to examine whether participants lost to follow-up were similar to those who completed the program and remained available for data collection. If not, then the results may not accurately represent the true effects of a program, which can often result in overly positive estimates.

Considerations regarding the impact of attrition on the validity of the program's outcomes are only one factor among several that are important in determining the overall validity of outcome data. It is also important to examine whether the program was implemented with fidelity and whether the program incorporated valid metrics to measure program processes and outcomes.

[Slide 62]

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 63]

On slide 63, what should I do if I am unable to obtain information about all of the financial aspects of my program?

Evaluators recognize that collecting accurate, highly detailed information about program costs can be quite burdensome. In general, we recommend collecting the most accurate and precise data needed to answer evaluation questions. Recall that accuracy is the degree to which measured values reflect the true value, whereas precision reflects the degree to which measured values are similar to one another.

This may not seem very intuitive, but if you must sacrifice data quality, then we suggest focusing on precision over accuracy. The reason for this preference is that precise measurement of costs

can be reproduced and compared over time.

Last, donated resources, such as labor provided by service members and civilian Defense Department employees, are a common feature of psychological health and traumatic brain injury programs. These resources are not generally reflected in program budgets and expenditures. However, it may be possible to produce estimates of these expenditures by using published pay scales containing average salaries by service branch or rank.

[Slide 64]

CAPT Thoumaian: Thank you, Dr. Sawyer, Dr. High, Dr. Hinds and Ms. Stark.

You've heard a great deal today about analyzing and interpreting data on program processes, outcomes and costs, 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 65]

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.

[Slide 66 to 68]

Ms. Stark: 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 66 to 68, we provide a brief list of key references and resources that we think may be useful.