

Program Evaluation Guide

Module 4: Developing a Data Plan



**DEFENSE CENTERS
OF EXCELLENCE**

For Psychological Health
& Traumatic Brain Injury

Table of Contents

Overview of the Program Evaluation Guide	1
Purpose and Use of the PEG	1
Purpose and Use of this Module	2
Determine Areas to be Measured.....	2
Begin by Examining Evaluation Questions.....	3
Review Program Logic Model and Objectives.....	3
Use Multiple Measurement Approaches Across Multiple Areas of Interest.....	3
Choosing, Refining or Developing Measures and Metrics.....	4
Validity and Reliability	4
How to Choose the Best Existing Measures and Metrics	5
Best Practices for Refining or Developing Measures and Metrics.....	5
Selecting Qualitative and Quantitative Methods	6
Qualitative Evaluation Methods	7
Quantitative Evaluation Methods	9
Mixed Methods	10
Data Entry and Storage	11
Preparing for Data Analysis	11
Develop a Data Matrix	11
Conclusion	13
Key Takeaways.....	13
References.....	14
Selected Resources for Additional Study.....	14
Appendix A. Data Matrix Examples.....	16
Template A. Data Matrix.....	17

Overview of the Program Evaluation Guide

This Program Evaluation Guide (PEG) is developed and published by the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury (DCoE). Program evaluation is an important part of the DCoE mission and helps military program administrators and leadership assess and improve service quality and outcomes. By making program evaluation an inherent part of everyday program activities, we create a culture of effectiveness to better build a sustainable, efficient and well-integrated continuum of prevention and care services for military members, their families and veterans.

The first edition of the PEG, published in July 2012, provided a standardized approach to program evaluation for psychological health and traumatic brain injury (TBI) program leaders. This version of the PEG (2nd Edition) has been updated and revised to reflect the most current needs of psychological health and TBI programs. This edition of the PEG is organized as a series of modules containing content specifically designed for use by program administrators or other staff members tasked with internal program evaluations as part of their duties within Defense Department psychological health and TBI programs. This PEG is designed for those who have limited prior knowledge and experience with the conduct of program evaluation activities.

Purpose and Use of the PEG

This PEG is one part of a collection of trainings, toolkits and support services offered by DCoE to assist personnel at the program level in developing their capabilities to conduct internal program evaluation activities. The PEG is designed for use in coordination with other training materials, such as the DCoE program evaluation and improvement webinar series, references provided in the PEG and webinar series, consultation with experts and other resources that may be available to program personnel.

The modules in this PEG are not intended to serve as a substitute for formal coursework on evaluation methods, statistics or data management. In addition, because the PEG is intended for use by a wide variety of programs, it will not provide specific guidance to programs on best practices for clinical or non-clinical services. Finally, the PEG is not intended as a manual for how evaluators who are external to a program should conduct their activities. However, the information herein will generally be useful in helping program personnel become more familiar with the evaluation process and consequently more effective in responding to external evaluation initiatives.

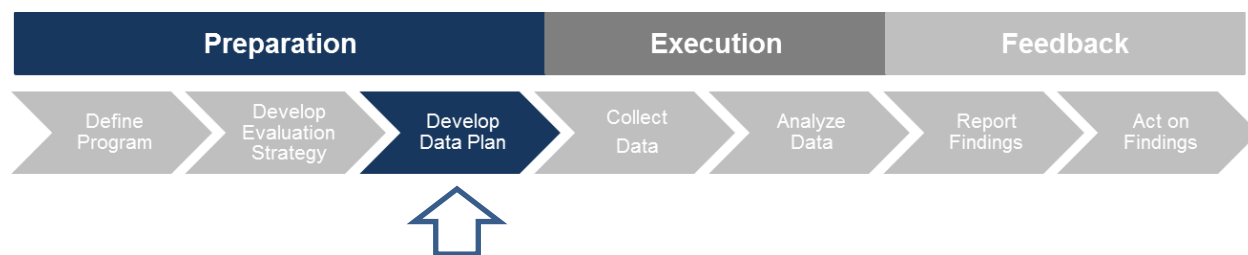
Developing a Data Plan

Purpose and Use of this Module

Once evaluation questions and strategies for addressing them have been developed, the program is ready to move to the next step of the evaluation process, Develop a Data Plan.

This module is designed to assist program personnel in preparing to execute the evaluation strategies identified in the previous step by developing a data plan. A data plan describes what is to be measured in an evaluation, how key areas of interest will be measured and how data are to be collected and stored. In addition, a data plan outlines how data will be analyzed following data collection and may include coverage of how data will be used in program improvement efforts or in ongoing program monitoring.

Because programs vary widely in their purposes and the populations they serve, this module provides broadly applicable guidance on developing a plan to execute an evaluation, while also indicating additional resources that may be used for selecting or developing metrics and conducting analyses tailored to each program.



Determine Areas to be Measured

Measurement is the process of collecting information, or data, about some area of interest. Often, people think of data in terms of numbers or quantitative data. However, data can be any type of information, including text, voice or video recordings, figures or drawings, and other non-numeric information that may be generated as part of an evaluation effort. A **metric** is the means by which data are collected, or the standard by which something is measured (Merriam-Webster, nd). Similarly, the term **measure** may be used to signify a specific data collection tool, such as a questionnaire or checklist.

The U.S. military has a long history of using measurement as an important part of its operations, including the Army Alpha and Beta test batteries, which were employed beginning in World War I to assign service members to duties based on their skills and abilities. Today, the Defense Department collects, stores and analyzes a wealth of data on service members through such efforts as the Armed Forces Health Surveillance Center, Defense Manpower and Data Center, Department of Defense (DoD) Suicide Event Report database and many other service- and program-specific initiatives.

In recent decades, there has been increased focus on demonstrating program effectiveness and maintaining accountability to stakeholders, taxpayers and funding agencies. Below, we discuss the process by which program personnel can determine areas should be measured.

Begin by Examining Evaluation Questions

The best place to begin to determine *what* should be measured is with evaluation questions developed as part of the previous step. Only after that step has been completed can program personnel determine *how* those areas of interest will be measured in terms of specific metrics or measures (Centers for Disease Control and Prevention [CDC], 2011a, 2011b). As an example, consider the evaluation question, “Is the program reaching its intended population?” This question requires that a program measure the targets of its service efforts (e.g., how many and what kinds of people participate). Similarly, the question, “Is the program achieving its intended outcomes?” demands that program personnel measure outcomes (e.g., changes in learning, skills, symptoms or functioning).

Review Program Logic Model and Objectives

The program logic model and objectives provide increasing specificity as to what will be measured and how it will be measured. Recall that a logic model includes four basic components: inputs, activities, outputs, and outcomes. A logic model specifies the core elements of a program and may include metrics, such as dollars spent across resource categories (i.e., an input), frequency and type of operations in which staff are engaged (i.e., activities), number and type of products and participants (i.e., outputs) and the amount of change that results in participants (i.e., outcome).

Provided that a program has a detailed logic model and SMART (specific, measurable, achievable, relevant and time-bound) objectives, deciding on measurement areas should be a relatively straightforward process. That is, selecting what to measure will simply involve matching evaluation questions with objectives and components of the logic model.

Use Multiple Measurement Approaches Across Multiple Areas of Interest

Measurement strategies are most effective when they use multiple methods and multiple sources of data (i.e., differing informants) within each area of interest. Rather than being redundant, multiple approaches can provide complementary or differing viewpoints. In addition, multiple approaches help the evaluator to compensate for the limitations of any single measurement strategy. Likewise, by conducting measurement across multiple areas, an evaluator can better capture the full breadth of how a program works and how it affects participants. By conducting measurement in multiple areas (e.g., both resiliency and learning, posttraumatic stress and depression, memory and pain), program personnel are better able to assess a program’s strengths and limitations, creating greater opportunities to both demonstrate effectiveness and to improve the program in targeted areas.

Figure 1 provides an example of three possible outcome areas that could be measured as part of a program evaluation effort. Each area includes two specific metrics, each with its own source of information. The *resiliency* area, for instance, is measured by providing a self-report questionnaire to program participants (e.g., service members completing a resiliency training program) and also by conducting focus groups with participants in which they discuss the relative benefits of the program in building resiliency. *Job functioning*, which is related to resiliency, is measured by days of work missed per year gathered from administrative records and also by brief interviews with unit commanders about participating service members’ performance on the job. Finally, the *family relationships* area is measured by questionnaires provided to participants and their spouses/partners as well as individual interviews with spouses/partners. In effect, these multiple measurement strategies reflect different angles of view to determine whether the program is affecting key areas.

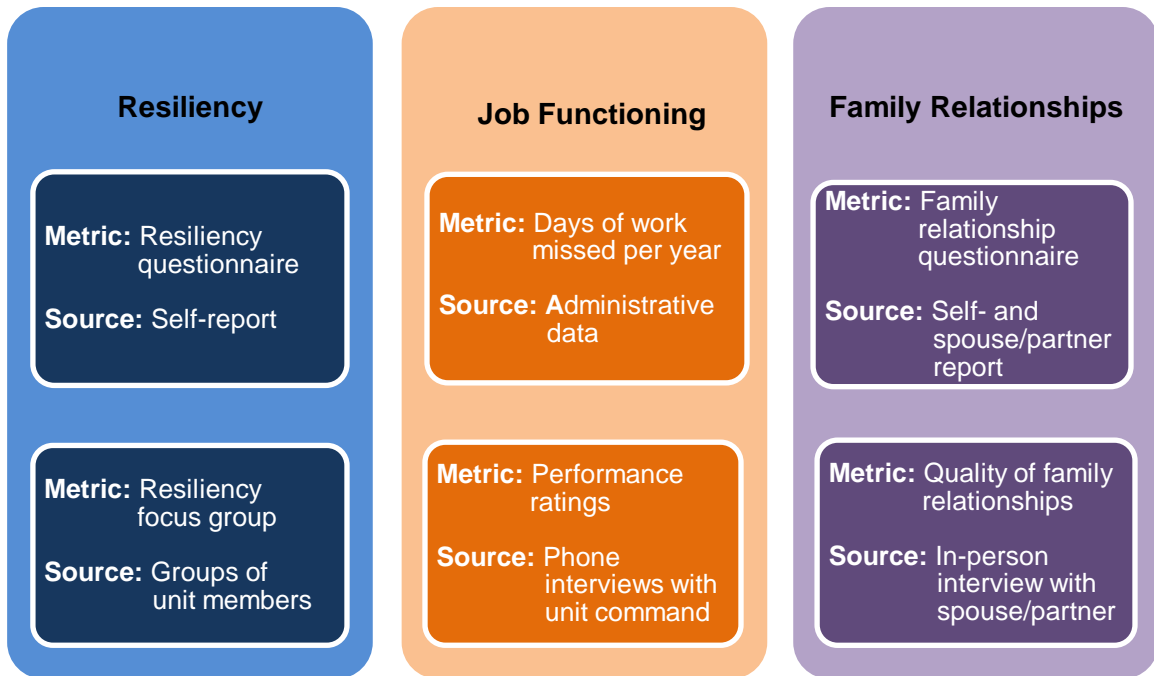


Figure 1: Sample Measurement Activities Across Three Outcome Areas

Multiple approaches may mean multiple quantitative approaches (e.g., different symptom or behavior scales to assess the nature of post-traumatic stress) or multiple qualitative approaches (e.g., an anonymous comment box as well as interviews with participants and staff) and combinations of both approaches. Of course, the need for rigor in evaluation must be balanced against practical constraints such as the time, energy and financial costs required to obtain and analyze the resulting data (Administration for Children and Families, 2010). There is a point of diminishing return when adding additional metrics; in general two to three metrics per area of interest are recommended when possible.

Choosing, Refining or Developing Measures and Metrics

Once program personnel decide what areas to measure, the next logical step is to decide how, specifically, to go about measuring them. This involves choosing or refining existing measures and metrics, or in some cases, developing a new measure or metric. Generally speaking, it is best to choose existing measures with proven validity and reliability when possible. There are numerous resources available to aid this process, such as those listed in the “Selected Resources for Additional Study” section at the end of this module.

Validity and Reliability

With respect to measurement, **validity** is the degree to which a measure or metric actually measures what it is intended to measure. **Reliability** refers to consistency of results, or the degree to which the same data would result from repeated measurement (Newcomer, Hatry, & Wholey, 2010). To use a sporting analogy, validity is equivalent to accuracy as indicated by how close a dart hits to the bullseye, the intended target. Reliability is equivalent to precision as indicated by how close together darts are (even if that means a tight cluster of darts stuck in the floor 10 feet in front of the dart board). It is not possible to achieve 100 percent validity and reliability; in other words, all measurement includes some degree of error or bias (Thorndike, 1997). However, program personnel should aspire to the highest

degree of measurement reliability and validity that are realistically possible. Ideal measures are both valid and reliable.

When choosing, refining or developing measures or metrics, it is important to acknowledge that validity and reliability are specific to a given purpose and population. In other words, a measure that has been shown to be valid in one context may not be equally valid or reliable in another context. For example, a learning assessment used for a web-based training presentation with senior military leadership may not be equally valid or reliable if used in a classroom-based training with junior enlisted personnel. The meaning of items may change due to varying experiences, education levels and the format in which information is presented, so it is important to be aware of possible sources of measurement error.



Ideal measures are both valid and reliable.

How to Choose the Best Existing Measures and Metrics

Whenever possible, standardized instruments that have demonstrated validity and reliability for the intended purpose and population should be used in lieu of new or modified measures. Using measures that are already validated has several advantages, most notably that the results can be compared to the outcomes reported from other programs and/or benchmarks established in the relevant research literature. When programs use non-standardized scales or questionnaires, the ability to generalize or compare findings is limited. In addition, validated measures lend greater credibility to evaluation findings and may help to dispel criticisms of evaluation findings.

Resources for finding existing measures include published lists of validated measures (see Selected Resources for Additional Study), relevant scientific literature and consultation with subject matter experts. As mentioned above, it is often beneficial to choose multiple measures or metrics for the same area of interest. Choices of measures should, however, include consideration of costs in terms of money, time, and the effort required to train users in administration. In addition, some measures may require that personnel obtain a user license and/or meet specific qualifications (e.g., hold a clinical license in a specific field, such as medicine or psychology). Finally, personnel involved in evaluation should assess whether suitable alternatives are available, alignment of measures with stakeholder interests, overlap with other measures used in the continuum of services and the practical usefulness of measures in guiding program services.

Best Practices for Refining or Developing Measures and Metrics

In some cases, it may be advisable to refine or modify an existing measure, such as when an existing measure is applied to a new population. For example, it may be necessary to adapt language for certain items in a measure designed for combat-exposed service members to better fit the experiences of a population that has not been directly exposed to combat, as some questions may not be applicable.

Likewise, personnel involved in program evaluation may wish to shorten an existing measure for use as a brief screening tool when a validated short-form of that measure is not readily available. It is important to know, however, that a shortened measure will often have lower reliability than its longer version, a phenomenon known as *shrinkage*.

New or custom measures may also be developed for novel purposes or populations. However, they are most likely to be used when collecting feedback from participants or assessing learning relevant to a specific program process (e.g., skills training). There are a number of best practices that may be used to ensure high quality measurement when developing items for a new or refined measure, as listed below in Table 1.

Table 1: Best Practices for Developing or Refining Measures

Check	Best Practice
<input type="checkbox"/>	Use a 6 th to 8 th grade reading level to accommodate varying language skills and cultural backgrounds
<input type="checkbox"/>	Provide participants with clear instructions on how to complete the measure
<input type="checkbox"/>	Train staff in how to administer measures, and conduct quality assurance checks on an ongoing basis
<input type="checkbox"/>	Avoid complex language, such as items with multiple clauses (i.e., “double-barreled” questions), metaphors or culture-specific examples
<input type="checkbox"/>	Spell out all acronyms and abbreviations the first time they are used
<input type="checkbox"/>	Use items that measure what they appear to measure (i.e., “face valid” items)
<input type="checkbox"/>	Be consistent in the number and type of response choices offered (e.g., 5-point response scale, True-False, Yes-No)
<input type="checkbox"/>	Obtain input on measure content from multiple sources, such as similar measures, experts, staff and program participants
<input type="checkbox"/>	Consult an expert if translating between languages, developing measures for individuals with limited language capabilities (e.g., children) or working with a population that includes individuals with hearing or vision impairments
<input type="checkbox"/>	Pilot test the measure before use

In an ideal world, modified and new measures would be subjected to formal validation studies. However, many programs will lack the resources needed to conduct a formal validation study of new or modified measures as such studies require substantial time and effort, large test populations and a high level of technical expertise. Program personnel may wish to consult with experts in measurement and evaluation when modifying or developing new measures. At a minimum, program personnel should seek to pilot test all measures and measurement procedures before they are used to collect data in an evaluation (Cook, Godiwalla, Brooks, Powers, & John, 2010). Pilot testing involves the use of role-plays and/or live practice with a subset of participants to ensure that procedures are clear and streamlined, that the procedures and measures are acceptable to staff and participants, and that administration is standardized so as to reduce variation that will result in unwanted error. In other words, practice measurement procedures before the evaluation period begins to ensure that the end results are as valid and reliable as possible.

Selecting Qualitative and Quantitative Methods

In all likelihood, an evaluation effort will include both quantitative and qualitative means of collecting and analyzing data. Most individuals are more familiar with quantitative approaches

than with qualitative approaches, although it is likely that program personnel have used qualitative methods extensively without having labeled them as such.

Qualitative Evaluation Methods

Qualitative methods involve the collection and analysis of non-numeric data (e.g., text, voice recordings) and can provide in-depth, highly contextualized information about a program, how it operates and/or how individuals such as staff and participants experience the program. Qualitative data collection generally results in much more data than quantitative data collection efforts, and the data require specialized coding and analysis strategies designed to identify important patterns and themes.

Some of the most commonly used qualitative data types are summarized in Table 2 and described in the text below to assist you in selecting methods for data collection, and additional information is available in the “Selected Resources for Additional Study” section.

Table 2: Common Methods of Qualitative Data Collection

Methods	Description	Characteristics
Interviews	One-on-one conversation	Can be structured or semi-structured
Focus groups	Group conversation facilitated by moderator	Use structured protocol with groupings of similar individuals
Open-ended comments	Written response on feedback forms or surveys	Voluntary expression of thoughts, opinions, suggestions
Observations	Log or description of activity	Applied in consistent manner to minimize bias
After action reviews	Group review following activity	Focus on strengths and opportunities for improvement
Case studies	In-depth longitudinal observations	Study of one individual, process or program over time

- **Interviews** involve a one-on-one conversation between an evaluator and a stakeholder, program administrator, provider of program services and/or a program participant. Interviews offer valuable insights into how a program actually functions in a private setting that may allow for more open expression of individual perspectives than in group settings. They yield detailed information, since the evaluator can ask follow-up questions. Interviews vary in their degree of structure, or how much the format of the interview constrains potential responses to interview questions and the ability of evaluators to pursue areas for follow-up. It is generally recommended that interviews be conducted with individuals representative of multiple groups in order to obtain a comprehensive view of the program.

- **Focus groups** use a moderator to guide a structured discussion among a small group of individuals. Current recommendations call for groups with as few as four and as many as a dozen people, although smaller groups are generally best (Krueger & Casey, 2010). The moderator should be someone with focus group experience and/or training. Rich information can be obtained when there is a free-flowing discussion among focus group participants. To that end, it is advisable that focus groups consist of individuals who are similar to one another in terms of characteristics like status or rank, so that group members do not feel overly inhibited in their responses. Both focus groups and interviews make use of a guide containing key questions and possible follow-up prompts related to the purpose of the evaluation. As with interviews, it is generally recommended that focus groups be conducted with multiple groups of interest to provide a comprehensive picture of the program.
- **Open-ended comments** are written responses to prompts, such as “Use this area to provide suggestions for improvements to this activity,” or simply the directive “Explain,” following a quantitative rating or yes/no response. The free-text response area allows respondents to write or type their thoughts, opinions, emotional reactions and suggestions. Responses are often anonymous in nature, although they may be collected from surveys or social media discussions (e.g., a blog, Facebook) in which respondents are identifiable. Because open-ended comments are voluntary, some caution should be exercised in interpreting responses, given that they may not represent the broader population of interest.
- **Observational techniques** are used to document activities, processes or conditions of a program or facility. Often a checklist is employed so that observers record what they see in a consistent manner and spend more time observing than writing. It is important that observers remain aware of privacy concerns when using this method and that they seek to minimize intrusion on the activity being observed, as the act of observation may cause those being observed to behave differently.
- **After action reviews**, also known as *hot washes*, are a type of qualitative process review in which the staff members who participated in an activity discuss its strengths and opportunities for improvement. They should result in an action plan and may be summarized in a written report. After action reviews are especially relevant to pilot testing of measurement processes.
- **Case studies** gather detailed information over time about a single entity, such as a participant or a program. For example, a single participant may be followed from program entry through program exit, or a program observer may document his or her experiences over the course of a year. Case studies are especially effective in “telling the story” of a participant or program in a detailed and compelling way.

Qualitative data are often considered to be “soft” data in that they are more open to interpretation and more difficult to replicate than quantitative, or “hard” data. However, qualitative data provide a high level of detail that give evaluators a more contextualized understanding of a program that cannot be achieved through quantitative means alone. Because stakeholders are generally less familiar with qualitative methods, we recommend having explicit conversations about their value to ensure support for their use. Likewise, it is essential that any personnel involved in qualitative data collection and analysis procedures receive appropriate training.

Quantitative Evaluation Methods

Quantitative methods involve the collection and analysis of numeric data (e.g., ratings, counts, frequencies, averages). Statistical analyses are used to identify patterns in quantitative data and draw conclusions that can often be applied beyond the immediate context in which the data were collected. Information about populations can easily be obtained and displayed, including demographic information (e.g., age, race, ethnicity, sex), program performance (e.g., number of services provided) and outcomes (e.g., changes in attitudes, skills, symptoms).

Three common sources of quantitative data are seen below in Table 3.

Table 3: Types of Quantitative Data

Source	Description	Characteristics
Questionnaires	Multi-item standardized measure that requires individual to provide response to questions or statements about some quality or characteristic	Can be administered in-person or via mail, telephone or electronically; may contain ratings (e.g., 1 to 5), yes/no, true/false, or frequency
Learning assessments	Multi-item test or quiz in which individual selects correct responses from a list of choices	Can be administered through multiple media; generally only one correct response
Structured screening protocols	One-on-one interview designed to generate specific information	Interviewer gathers specific information from individual and provides ratings based on predetermined scoring rules

The source column contains three commonly-used quantitative data collection methods. Questionnaires and learning assessments are both multi-item forms completed on paper or using a computerized format.

Questionnaires ask for responses about some quality or characteristic, like quality of life or symptoms of TBI. Learning assessments are tests or quizzes in which responses indicate the degree of learning or knowledge; they are commonly used to examine the results of training programs.

Structured screening protocols are like questionnaires delivered in an extended interview format. The interviewer gathers specific information that can be scored using a set of rules to generate quantitative data, like a diagnosis or a symptom total.

Quantitative data are collected by means of a variety of specific measures or metrics, including the following examples:

- Self-ratings on questionnaires and surveys
- Ratings by a commander, family member or clinician

- Learning and performance measures in terms of completion time or number of correct responses
- Body temperature, blood pressure or heart rate
- Checklist totals from observations or interviews

The resulting quantitative data may need to be scored, which could involve calculating a total or using a more complex scoring protocol to re-code certain items so that all responses align in the same direction (e.g., so that all positive responses mean the same thing). Some measures have associated scoring software capable of generating reports and rapidly checking for data entry errors and missing responses. However, program personnel can often create simple software programs of their own using Microsoft® Excel or other readily available software designed for data entry and storage, scoring, analysis and quality assurance (e.g., IBM's® Statistical Package for the Social Sciences).

Mixed Methods

There are strong practical reasons to view qualitative evaluation methods as complementary to quantitative methods. Indeed, using mixed methods can often be better than just one method alone. There is no single measurement strategy that is best in all cases. Choices about measurement strategies should be tailored to meet the specific needs of an evaluation effort, even though the same general types of information are often important, such as documenting a program's use of resources, how it performs activities, the outputs of those activities and outcomes of participants.

Once evaluation questions are identified, evaluators must make important decisions as part of the planning process about the order in which qualitative and quantitative data will be collected and analyzed. Below are three designs that vary in the order and uses of qualitative and quantitative evaluation methods.

- In a **parallel design**, quantitative and qualitative data are collected simultaneously and data are merged and compared side by side during the analysis phase. Parallel designs are useful when mixed methods are used to answer the same evaluation question. For example, a feedback form containing both quantitative and qualitative components may be administered as a single unit at the same time.
- In a **sequential design**, differing methods are used separately with one before another. The results from the first method inform the design and execution of a second method (or more as needed). For example, an evaluator might conduct surveys, analyze the results of that survey, and then return to the program participants to conduct a qualitative focus group to learn more detailed accounts that help to explain the survey results or address additional follow-up questions that were not adequately addressed by the survey.
- Finally, **embedded or nested designs** use both qualitative and quantitative methods simultaneously to answer two or more distinct evaluation questions. For example, an evaluator might use quantitative methods (e.g., rating scales, questionnaires) to examine questions regarding whether a program is achieving its intended outcomes, while using qualitative interviews to assess staff and participant perceptions about the quality of services provided and areas for potential process improvements.

Data Entry and Storage

Regardless of the type of data collected, effective data entry and storage are important considerations. These processes are certainly among the least glamorous in the conduct of program evaluation; however, they ultimately determine the accuracy of the data collected and ensure that data are available for future use. Moreover, these are two areas in which errors and security breaches often occur due to careless and easily avoidable mistakes. At the outset of data collection, be sure to specify how data will be entered into a database and/or filed for storage (e.g., electronically, paper, disc), the location(s) in which they will be stored, and who will be responsible for storing and maintaining data and ensuring data security. Finally, regular quality assurance checks for accuracy of data entry and proper maintenance of records should be built into evaluation procedures if they are not already a part of a program's standard operating procedures (CDC, 2011b). Specific details regarding how to collect, enter and store data will be covered in a future PEG module.

Preparing for Data Analysis

When developing a data plan, it is always best to begin with the end goal in mind: to extract useful information from the data collected. Determine in advance how data will be analyzed and interpreted. Failure to engage in advanced planning for data analysis results in a problem commonly known as DRIP (Data Rich but Information Poor). Thus, it is important to anticipate the skills needed to analyze and interpret data, who will be assigned this task, what resources (e.g., time, software, funding, outside consultations) are needed and whether program personnel will need additional training. In addition, it is important to ensure that data entry and storage are conducted in a manner that allows analysis to occur efficiently. For instance, if scoring and recoding of data are needed, it is best to perform that activity at the data entry stage, rather than waiting until just before analyses are conducted. Specific analysis strategies will be described in future PEG modules. Finally, and perhaps most importantly, it is essential that program personnel collect the type of data needed to answer evaluation questions and to begin collecting those data if they are not already being collected. Prime examples include outcome and cost information, which immature programs (i.e., those that have only been in existence for a short time) may not be collecting but which are necessary to answer common evaluation questions such as, "Is the program producing its intended outcomes?" and "How much does the program cost per participant?"

Develop a Data Matrix

A data matrix is a useful way of summarizing data collection, entry, storage and analysis plans as part of any larger data plan. In the example presented below in Table 4, consider a non-clinical program that provides training to service members (e.g., on resiliency skills or stress management). To address the evaluation question, "Is the program achieving its intended outcomes?", three different metrics are used to measure the effects of the program on participant outcomes, namely whether and how much participants learn from the program activities delivered by the event's organizer and trainers.

- The **Learning Metric 1** column indicates that the training event organizer will collect post-test quizzes from event participants. The organizer is also responsible for entering the data into a storage system (e.g., a database), and the program manager will conduct data analysis and interpretation in order to track outcomes and write reports.
- The **Learning Metric 2** column provides similar details for ratings collected in the web-based feedback forms sent out immediately after the event and gathered by the program's website manager/IT support staff. Feedback may be considered an output or

outcome and is used for tracking purposes, to provide feedback to trainers and in reporting how participants perceive the training.

- The **Learning Metric 3** column indicates that a random subset of participants will be called three months after the event to determine whether and how the information provided at the training he or she received was actually put into practice. This qualitative information may provide the program with useful information about its longer-term effects and is analyzed by the program manager for the purposes of reporting to stakeholders as well as ensuring the program is continuously monitoring its quality and making improvements to refine its training practices.

Appendix A presents two additional examples of data matrices, and a blank template is provided below in Template A.

Table 4: Sample Data Matrix for a Non-Clinical Training Program

Matrix Item	Learning Metric 1	Learning Metric 2	Learning Metric 3
Data source	Participants	Participants	Select participants
Data collection method	Post-test quiz results	Anonymous feedback form ratings on items “Did training meet its objectives?” and “As a result of this training, my knowledge increased.”	Brief qualitative phone interviews with subset of randomly selected participants to determine how information was used
Who will collect data?	Training event organizer	Training event organizer	Training event organizer
Frequency	Post-event	Post-event	3-months post-event
Who will input data?	Event provider	Website manager	Event provider
Who will analyze data?	Program manager	Program manager	Program manager
How will data be used?	Outcome tracking, reporting	Output and outcome tracking, staff feedback, reporting	Reporting, continuous program improvement

The data matrix specifies key details of the next phase of the evaluation process – the Execution Phase – in a detailed manner. A broader data plan might include several data matrices, based on the number of evaluation questions and the number of areas to be measured for each question (e.g., different types of outcomes each with two to three metrics). In addition a broader data plan may include standard operating procedures, information about rules and regulations and other information central to carrying out an evaluation effort. A good data plan should anticipate ongoing data collection as part of continuous monitoring and quality improvement processes embedded in a program’s operational structure, because one-time-only evaluations are likely to be of limited value by themselves.

Conclusion

At the conclusion of this module, “Developing a Data Plan,” program personnel should have determined which areas to measure and begun to select or develop the specific qualitative and/or quantitative measures and metrics that will be used to evaluate their program. In addition, personnel involved in an evaluation should have laid out the details of how data will be entered, stored and maintained. Moreover, the data plan must anticipate how data will be analyzed and interpreted. Finally, a data matrix should have been developed to serve as a detailed framework featuring key details of the data plan for each evaluation question, such as who will carry out data collection and analysis activities and when these activities will be conducted.

Key Takeaways

- Choose areas to measure that match the evaluation questions designed for the program
- Use existing, validated measures when possible and engage in best practices if developing new measures or if refining or adapting previously existing measures
- Employ multiple measurement strategies to ensure that evaluation efforts provide a comprehensive picture of a program and to help overcome the limitations of any single measurement strategy
- Develop a detailed data matrix for each evaluation question as part of a larger data plan before executing data collection and analysis activities

References

- Administration for Children and Families, Office of Planning, Research and Evaluation (2010). *The program manager's guide to evaluation* (2nd ed.). Retrieved from U.S. Department of Health and Human Services website: <http://www.acf.hhs.gov/programs/opre/resource/the-program-managers-guide-to-evaluation-second-edition>
- Centers for Disease Control and Prevention (2011a). *Introduction to program evaluation for public health programs: A self-study guide*. Retrieved from: <http://www.cdc.gov/eval/guide/>
- Centers for Disease Control and Prevention (2011b). Developing an effective evaluation plan. Retrieved from: <http://www.cdc.gov/obesity/downloads/CDC-Evaluation-Workbook-508.pdf>
- Cook, S. C., Godiwalla, S., Brooks, K. S., Powers, C. V., & John, P. (2010) Recruitment and retention of study participants. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.) *Handbook of practical program evaluation* (3rd ed., pp. 182-207). San Francisco: Jossey-Bass.
- Krueger, R. A., & Casey, M. A. (2010) Focus group interviewing. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.) *Handbook of practical program evaluation* (3rd ed., pp. 378-403). San Francisco: Jossey-Bass.
- Merriam-Webster. (n.d.). *Merriam-Webster's online dictionary*. Retrieved from: <http://www.merriam-webster.com/>
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2010) Planning and designing useful evaluations. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.) *Handbook of practical program evaluation* (3rd ed., pp. 1-29). San Francisco: Jossey-Bass.
- Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (6th ed.) Upper Saddle River, NJ: Prentice Hall.

Selected Resources for Additional Study

- Agency for Healthcare Research and Quality: <http://www.qualitymeasures.ahrq.gov>
- American Evaluation Association, Instrument Collections: <http://www.eval.org/p/cm/ld/fid=72>
- Center for Quality Assessment and Improvement in Mental Health, National Inventory of Mental Health Quality Measures: <http://www.cqaimh.org/NIMHQM.htm>
- DCoE Program Evaluation Guide and Program Evaluation Webinar Series: http://www.dcoe.mil/About_DCoE/Program_Evaluation/Resources_and_Training.aspx
- Deployment Health Clinical Center, Assessment Tools: www.pdhealth.mil/clinicians/assessment_tools.asp
- Defense and Veterans Brain Injury Center, Diagnosis and Assessment: [http://dvbic.dcoe.mil/diagnosis-assessment?audience\[0\]=3](http://dvbic.dcoe.mil/diagnosis-assessment?audience[0]=3)
- Institute of Medicine Reports:
For the Public's Health: The Role of Measurement in Action and Accountability (2010): <http://www.iom.edu/reports/2010/for-the-publics-health-the-role-of-measurement-in-action-and-accountability.aspx>

Preventing Psychological Disorders in Service Members and Their Families: An Assessment of Programs (2014): www.iom.edu/Reports/2014/Preventing-Psychological-Disorders-in-Service-Members-and-Their-Families.aspx

Treatment for Posttraumatic Stress Disorder in Military and Veteran Populations: Final Assessment (2014): <http://www.iom.edu/Reports/2014/Treatment-for-Posttraumatic-Stress-Disorder-in-Military-and-Veteran-Populations-Final-Assessment.aspx>

Joint Commission on Accreditation in Health Care, Hospital Based Inpatient Psychiatric Services Core Measurement Set: <http://www.jointcommission.org/assets/1/6/HBIPS.pdf>

Medical Outcomes Trust, Instruments: <http://www.outcomes-trust.org/instruments.htm>

Buros Center for Testing, Mental Measurements Yearbook: <http://buros.org/mental-measurements-yearbook>

Military Health System, 2014 Report: Final Report to the Secretary of Defense Military Health System Review: <http://www.health.mil/Military-Health-Topics/Access-Cost-Quality-and-Safety/MHS-Review>

National Center for PTSD, Assessment Overview: <http://www.ptsd.va.gov/professional/assessment/overview/index.asp>

National Institutes of Health Toolbox: <http://www.nihtoolbox.org/>

National Institute of Neurological Disorders and Stroke, Common Data Elements: <http://www.commondataelements.ninds.nih.gov/>

National Quality Forum, Measures, Reports, and Tools: www.qualityforum.org/Measures_Reports_Tools.aspx

Society for Prevention Research, Standards of Evidence: <http://www.preventionresearch.org/advocacy/#SofE>

Substance Abuse and Mental Health Services Administration, National Behavioral Health Quality Framework: <http://www.samhsa.gov/data/national-behavioral-health-quality-framework>

The Center for Outcome Measurement in Brain Injury: <http://www.tbims.org/combi/list.html>

U.S. Department of Health and Human Services, Health Information Privacy: <http://www.hhs.gov/hipaa/index.html>

U.S. Department of Veterans Affairs, Health Services Research & Development: <http://www.hsrp.research.va.gov/>

Veteran's Health Administration Mental Health Program Evaluation Technical Manual: http://www.rand.org/pubs/working_papers/WR682.html

Appendix A. Data Matrix Examples

The examples below provided completed data matrices for a hypothetical clinical program and non-clinical program, respectively.

Example 1 Evaluation Question (Clinical Program): Is the program achieving its intended outcomes?

Matrix Item	Outcome Metric 1	Outcome Metric 2	Outcome Metric 3
Data source	Participant	Provider	Commander
Data collection method	Self-report questionnaire on symptoms	Functional rating across work, family and individual areas	Functional rating of work performance
Who will collect data?	Service provider	Service provider	Program manager
Frequency	Pre- and post-treatment, 3-month follow-up	Pre- and post-treatment, 3-month follow-up	Pre- and post-treatment
Who will input data?	Service provider	Service provider	Program manager
Who will analyze data?	Program manager	Program manager	Program manager
How will data be used?	Treatment planning and outcome tracking	Outcome tracking and reports	Outcome tracking and reports

Example 2 Evaluation Question (Non-Clinical Program): Is the program reaching its intended audience?

Matrix Item	Participation Metric 1	Participation Metric 2
Data source	Registration page	Website
Data collection method	Attendance on webinar day (number of attendees, demographics)	Document downloads and web page visits
Who will collect data?	Event organizer	Event organizer
Frequency	On event day	Monthly
Who will input data?	Event provider	Website manager
Who will analyze data?	Program manager	Program manager
How will data be used?	Output tracking and reports	Output tracking and reports

Template A. Data Matrix

Use the template below to specify details of a data collection plan for each major evaluation question. Program personnel may wish to modify this template by including additional information, such as resources and training required, special considerations regarding privacy and data security and/or rules and regulations that apply to data collection and storage.

Evaluation Question: [Insert text]

Matrix Item	Metric 1	Metric 2	Metric 3
Data source			
Data collection method			
Who will collect data?			
Frequency			
Who will input data?			
Who will analyze data?			
How will data be used?			