

PROGRAM EVALUATION GUIDE





DEFENSE CENTERS OF EXCELLENCE
For Psychological Health & Traumatic Brain Injury

Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury

Program Evaluation Guide

July 2012

2345 Crystal Drive ★ Crystal Park 4, Suite 120 ★ Arlington, Virginia 22202 ★ 877-291-3263

1335 East West Highway ★ 9th Floor, Suite 640 ★ Silver Spring, Maryland 20910 ★ 301-295-3257

www.dcoe.health.mil ★ Outreach Center: 866-966-1020

TABLE OF CONTENTS

Introduction	4
Overview of the Program Evaluation Guide	4
How to Use this Guide	4
Background	5
Program Evaluation Overview	6
Program Evaluation Definition	6
The Benefits of Program Evaluation	6
Types of Program Evaluation	7
Why and When to Conduct Program Evaluations	9
Distinctions Between Program Evaluation, Performance Improvement and Formal Research	10
Reasons for Initiating Program Evaluations	11
Team Composition	11
A Program Evaluation Framework	12
Key Takeaways	14
Program Evaluation Methodology: Phases and Steps	16
Phase 1: Preparation	16
Step 1: Review of Program	17
Step 2: Develop Evaluation Questions	31
Step 3: Develop Evaluation Design	42
Develop the Evaluation Design	42
Understanding Internal Validity	46
Step 4: Develop a Data Plan	49
Step 5: Gather Data	57
Step 6: Analyze Data	60
Step 7: Develop Report	72
Step 8: Act on Findings	76
Conclusion	78
Appendices	79
Appendix I: Examples of Program Evaluation (PE)	80
Appendix II: Introduction to Cost Analysis for Program Managers	85
Appendix III: Official Instructions, Policies, Regulations and Guidance Related to Program Evaluation	91
Appendix IV: Guide to Conducting Interviews and Focus Groups for Program Evaluation ..	93
Interviews	93
Focus Groups	95
Appendix V: Sources of Effectiveness Measures	100
Appendix VI: Information on Recommended Measures for PH and TBI	101
Appendix VII: Conducting a Follow-Up Evaluation	103
Appendix VIII: Guide to Coding Qualitative Data	106

Appendix IX: Overview of Basic Statistical Analyses 109
 Appendix X: Templates 112
 Appendix XI: Frequently Asked Questions 113
 Appendix XII: Glossary of Terms 118

FIGURES

Figure 1. The Impact of Program Evaluation 7
 Figure 2. Evaluation Types and Levels of Evidence 7
 Figure 3. Selecting a Type of Program Evaluation 9
 Figure 4. Relationship Between Program Evaluation, Performance Improvement and
 Formal Research..... 10
 Figure 5. Evaluation Team Roles and Responsibilities..... 11
 Figure 6. DCoE Program Evaluation Framework..... 12
 Figure 7. Example Program Evaluation Timeline..... 14
 Figure 8. Examples of Mission Statements..... 19
 Figure 9. Relationship Between Mission, Goals and Objectives..... 19
 Figure 10. Developing Program Goals..... 20
 Figure 11. The SMART Framework 21
 Figure 12. Example Objectives 22
 Figure 13. Logic Model Components 24
 Figure 14. Example Logic Model for a Specialty Clinic for Combat-Related PTSD..... 28
 Figure 15. Example Logic Model for a Specialty TBI Clinic 29
 Figure 16. Example Logic Model for a Substance Abuse Clinic 30
 Figure 17. Scope of Analysis for Process vs. Outcome Evaluations 32
 Figure 18. Sample Evaluation Questions by Evaluation Type 33
 Figure 19. Sample Evaluation Questions 34
 Figure 20. Common Data Collection Tools 35
 Figure 21. Sample Output Measures for Process Evaluations 37
 Figure 22. Sample Outcome Measures for Outcome Evaluations 37
 Figure 23. PH and TBI Instruments 38
 Figure 24. Types of Reliability and Validity 40
 Figure 25. Threats to Internal Validity 47
 Figure 26. Levels of Data Measurement..... 62
 Figure 27. Frequency Table for Patient Category Data 63
 Figure 28. Bar Chart for Patient Category Data 63
 Figure 29. Pie Chart for Patient Category Data 64
 Figure 30. Inferential Statistics Criteria 66
 Figure 31. Participation, Outcome, and Effectiveness 69
 Figure 32. Cost Analysis Logic Model..... 85
 Figure 33. Steps in Cost Analysis 86
 Figure 34. Pros and Cons for Data Gathering Methods..... 104
 Figure 35. Pros and Cons for Contact Methods..... 105
 Figure 36. Steps to Coding Qualitative Data for a Program Evaluation 106
 Figure 37. Example of Coding Qualitative Data 108
 Figure 38. Sample Research Questions 109
 Figure 39. Descriptive and Inferential Statistics..... 109
 Figure 40. Sample Statistical Tests 110

Figure 41. Statistical Options Flowchart 110
 Figure 42. Sampling Methods 111
 Figure 43. DCoE Program Evaluation Framework..... 114

ACRONYMS

AHLTA	Armed Forces Health Longitudinal Technology Application
ANOVA	Analysis of variance
ANCOVA	Analysis of covariance
AUDIT	Alcohol Use Disorders Identification Test
AUDIT-C	AUDIT alcohol consumption questions
CAPS	Clinician-Administered PTSD Scale
CHCS	Composite Health Care System
CT	Computed tomography
DCoE	Defense Centers of Excellence
DHI	Dizziness Handicap Inventory
DoD	Department of Defense
DVBIC	Defense and Veterans Head Injury Center
EHR	Electronic health records
FHP&R	Force Health Protection and Readiness
HA	Health Affairs
HIPAA	Health Insurance Portability and Accountability Act of 1996
ImPACT	Immediate Post-Concussion Assessment and Cognitive Testing
IRB	Institutional Review Board
LIMDU	Limited duty
MACE	Military acute concussion evaluation
MDD	Major depressive disorder
MEB	Medical Evaluation Board
MHS	Military Health System
MRI	Magnetic resonance imaging
mTBI	Mild traumatic brain injury
NICoE	National Intrepid Center of Excellence
NSI	Neurobehavioral Symptom Inventory
OT	Occupational therapy
PCL	PTSD check list
PCL-M	PTSD check list – military version
PE	Program evaluation
PH	Psychological health
PHQ-9	Patient Health Questionnaire
PII	Personally identifiable information
PTSD	Post-traumatic stress disorder
SLP	Speech language pathology
SME	Subject matter expert
SOP	Standard Operating Procedures
SUD	Substance use disorder
TBI	Traumatic brain injury
TMA	Tricare Management Activity
VA	U.S. Department of Veterans Affairs

Introduction

OVERVIEW OF THE PROGRAM EVALUATION GUIDE

In 2010, at the request of Force Health Protection and Readiness (FHP&R), the Defense Centers of Excellence (DCoE) for psychological health (PH) and traumatic brain injury (TBI) worked with program and portfolio leaders of several Department of Defense (DoD) PH and TBI pilot and demonstration projects to evaluate their programs' effectiveness. These evaluations revealed a common theme regarding the programs. While program leaders understood that monitoring and evaluating effectiveness provide a valuable means to improve service delivery, they lacked structured resources to prepare for and conduct methodological program evaluation practically and efficiently.

Under direction from FHP&R, DCoE tailored its standardized program evaluation methodology for PH and TBI for program leader use. The DCoE *Program Evaluation Guide* (subsequently referred to as the *Guide*) provides a step-by-step "how to" manual for program managers to prepare for and conduct program evaluations. The program evaluation (PE) methodology provides PH and TBI program managers a consistent and analytic approach to evaluate the impact of their programs. The methodology in the Guide could also be used by an outside evaluator, or by program staff to conduct a self-evaluation. The Guide is organized as a resource that can be used to plan the evaluation as well as a reference that can be returned to as each step is executed. This framework consists of three phases which link together eight steps.

HOW TO USE THIS GUIDE

Throughout this Guide, detailed steps and templates are provided. Whenever appropriate, templates have been provided to structure key activities. These templates were developed for program managers to reduce the time and effort required to plan, execute and report on the findings of the program evaluation. To maximize convenience, the templates referenced throughout the document are all embedded in the Guide as figures in the section where they are introduced. **To view and utilize the templates, simply double-click on the picture of the template**, which will open as a separate document in another window. To print the document so that it can be filled in by hand, send directly to print. To type in the content of the template fields, **save the template document first**. At the end of each section, there is a listing of sources used in developing that section's content and suggested readings for topics presented in that section.

Background

The Defense Centers of Excellence (DCoE) for Psychological Health (PH) and Traumatic Brain Injury (TBI) is charged with identifying programs that positively impact care to service members and their families. DCoE was established in 2007 with a 900 million dollar Congressional appropriation. These funds are allocated for the integration of DoD and U.S. Department of Veterans Affairs (VA) research and treatment initiatives that address TBI and PH problems in service members. Although TBI and psychological health problems, such as post-traumatic stress disorder (PTSD), were present in previous military conflicts, the nature of combat in the wars in Afghanistan and Iraq has resulted in significantly greater attention and resources devoted to these issues. The increased prevalence and visibility of these issues has prompted the establishment of various DoD programs to address the quality of service and overall care provided to veterans, service members and their families.

Since the charter of DCoE in 2007, the number of programs available to serve service members with TBI and/or PH issues has grown significantly. As the number of programs has increased, so has the need for the DoD to minimize any duplication and redundancy of effort and to maximize the benefits and services for service members and veterans. The need for coordination and oversight of PH and TBI activities is especially critical given the projected five-year DoD budget reduction plan. In 2010, DCoE was tasked by FHP&R to conduct evaluations of the effectiveness of several DoD programs. DCoE concluded that while most programs were tracking some measure of program efficacy, additional support was required to establish sufficient measures demonstrating both statistically and clinically significant impact.

Program evaluation is a process that formalizes and provides a consistent structure for the review and assessment of a program. Leaders use a series of tools to assess effectiveness and scalability from both a qualitative and quantitative perspective, highlighting the programs' strengths, opportunities and lessons learned. Successful demonstration of the need, effectiveness and scalability of TBI and PH programs has the potential to facilitate program improvements and therefore provide increased value to program beneficiaries and stakeholders. The findings from formalized program evaluations strengthen not only the evaluated program itself, but may also be translated into benefits for other programs and efforts.

In an effort to strengthen the existing pool of program review and evaluation resources and tools, DCoE has developed this Program Evaluation Guide. The Guide, a step-by-step manual for program and portfolio leaders, provides information and instructions on methodical, practical and efficient program evaluation methods and techniques. It is designed to bolster a program leader's ability to develop or refine goals, establish effectiveness measures, and synthesize data to determine if a program is meeting its stated objectives. As a program evaluation roadmap, the Guide provides templates to examine program effectiveness and provide guidance on implementing programmatic changes to improve outcomes and meet specific goals. It is designed to aid program leaders in demonstrating statistically and clinically significant results that benefit service members and their families.

Program Evaluation Overview

PROGRAM EVALUATION DEFINITION

“Program evaluations are individual systematic studies conducted periodically or on an ad hoc basis to assess how well a program is working” (GAO, 2011).

The process involves the collection, analysis and interpretation of data to determine the outcomes and effectiveness of a program, adherence to mission, and identification of areas in need of improvement, as well as opportunities for growth.

Program Evaluation

- Often conducted by experts external to the program, drawn from the agency itself or from an external agency. Evaluations may also be conducted by program managers
- Typically examines achievement of program objectives in the context of other aspects of program performance.

(GAO, 2011)

THE BENEFITS OF PROGRAM EVALUATION

Program evaluations serve many important functions. Program evaluations can determine how well a program is working by measuring the achievement of the program’s objectives pertaining to aspects of program performance. Evaluating results will enable the program to further improve program performance and demonstrate impact. Many specific benefits of program evaluation are detailed in the box to the right. It should be noted that program evaluations are an iterative process, interacting with the program in an ongoing cycle. The figure below shows how a program evaluation yields results, which may have effects on how the program is structured (e.g., increasing staffing) or functions (e.g., implementing a process improvement project to more closely follow clinical practice guidelines). See **Figure 1. The Impact of Program Evaluation** below.

Benefits of Program Evaluation

- Identifying and advancing best practices
- Reducing redundancy
- Improving cost effectiveness
- Providing a feedback mechanism in a resource-constrained environment
- Developing a clear, common understanding of program to help with program management, improvement and both internal and external communication about the program
- Identifying program strengths
- Identifying specific opportunities for improvement, growth and development
- Understanding the unintended consequences of the program (both positive and negative)
- Establishing the program as an evidence-based practice
- Establishing data that demonstrate the program is (or is not) effective and understand why (or why not)
- Making a case for the program (to staff, funders, clients, etc.)
- Developing the capacity to conduct ongoing internal evaluations or undergo external evaluation

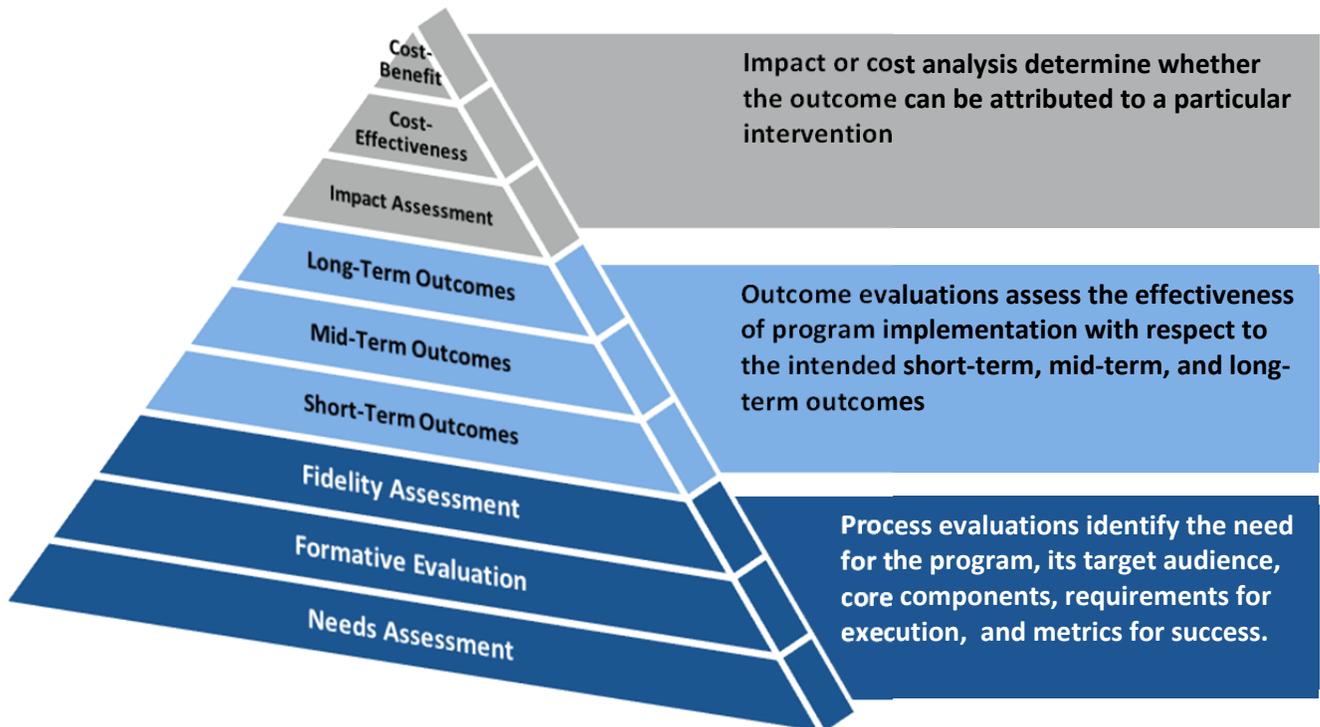
Figure 1. The Impact of Program Evaluation



TYPES OF PROGRAM EVALUATION

While there are many typologies for program evaluation, this Guide will discuss three types of evaluation: Process Evaluation; Outcome Evaluation and Impact Evaluation/Cost Analysis.

Figure 2. Evaluation Types and Levels of Evidence



Process Evaluation. A process evaluation is designed to assess the structure and activities of the program to ensure that it is operating as intended. Program inputs and resources, often in the form of time, effort and financial investments, are examined and measured against services delivered by the program. This type of evaluation can also help identify problems with how a program is being implemented and may suggest strategies for correction. Process evaluations are typically conducted early in a program's lifecycle, and it may be beneficial for programs to conduct many of these types of evaluations in the first years of operation. An example process evaluation concern could be: How well is the new TBI specialty clinic functioning? This evaluation could look at staffing levels, wait times for appointments, patient satisfaction, number of referrals, and other variables related to how the program is functioning. The subcategories of process evaluation include:

- **Needs Assessments:** Determine who needs the program, how great the need is, and establish goals and objectives of a program based on a systematic review and analysis of the target population's needs.
- **Formative Evaluations:** Validate that the goals of the program instructions, interventions, or activities are being achieved through use of surveys, focus groups, or review of process metrics.
- **Fidelity Assessments:** Determine whether a program's activities are being implemented in a manner consistent with the original intention. Fidelity assessments allow program managers to have more confidence that the changes in outcomes are actually due to the interventions (e.g., a PTSD treatment program using prolonged exposure therapy conducts a review of how closely providers are following the treatment protocols).

Outcome Evaluation. While process evaluations focus on how a program was/is being implemented, an outcome evaluation analyzes whether the program is effective. Outcome evaluations are typically not conducted on new programs, as it may take many months or even a period of a few years for a program to be able to show impact. An example of an outcome evaluation could be: Are the patients who have had treatment in the new TBI specialty clinic improving? The outcome evaluation would examine many of the same things as a process evaluation, but it would be specifically focused on outcome measures, for instance, "return-to-duty rate" or "improved symptom reporting after TBI." There are three subcategories of outcome evaluations:

- **Short-Term Outcomes:** Refer to the outcomes that expected directly following program activities. Typically these types of evaluations focus on participant satisfaction, and changes in knowledge, skills and attitudes.
- **Mid-Term Outcomes:** Refer to the outcomes expected some time after program implementation, and usually focus on changes in behavior, which might include declines in symptoms or risk-taking behaviors.
- **Long-Term Outcomes:** Refer to the larger community-based impact seen long after program implementation, which might include life satisfaction, disability rates, etc.

Impact Evaluation & Cost Analyses: Impact evaluations and cost analyses aim to identify what aspect of a program, policy, or process is responsible for an outcome. These types of evaluations examine the relationship between the inputs (resources put into a program) and the outputs or outcomes of the program. Like outcome evaluations, which also rely on having

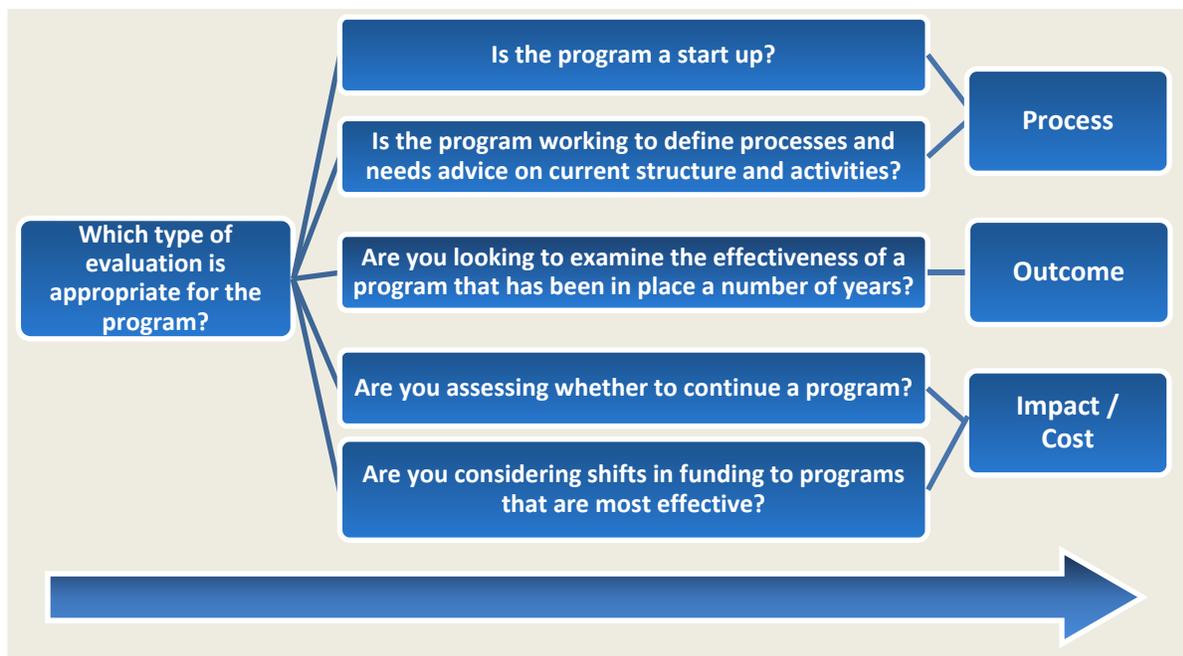
enough time for a program to have meaningful data on its effectiveness, these evaluations are usually conducted later in the lifecycle of a program. Three subcategories of impact evaluation include:

- **Impact Evaluations:** Determine whether the program itself is causing the observed outcomes or whether the outcomes are due to external factors. This type of analysis typically involves randomized controlled trials.
- **Cost-Effectiveness Analyses:** Relate the costs of a program to key outcomes or program benefits.
- **Cost-Benefit Analyses:** Relate the costs of a program to the monetary value of program benefits. A more detailed discussion of cost-benefit and cost-effectiveness analyses is provided in **Appendix II: Introduction to Cost Analysis for Program Managers.**

WHY AND WHEN TO CONDUCT PROGRAM EVALUATIONS

Program evaluation should be a fundamental component of program planning and implementation. The process of re-evaluating, revising and refining program activities on a regular basis is vital to ensuring the program is progressing and improving as well as meeting its mission. Thus it must be understood that program evaluation is an iterative process. That said, process evaluations may be more applicable to a startup that is still working to define its processes and could benefit from feedback on its current structure and activities. Outcome evaluations might benefit a program that has been in place for a number of years to examine whether the program is effective. Lastly, impact and cost-benefit analyses may be appropriate when assessing whether to continue with a particular program or when a change is being considered. **Figure 3. Selecting a Type of Program Evaluation** shows decision tree to assist in selecting the most appropriate type of evaluation for the situation.

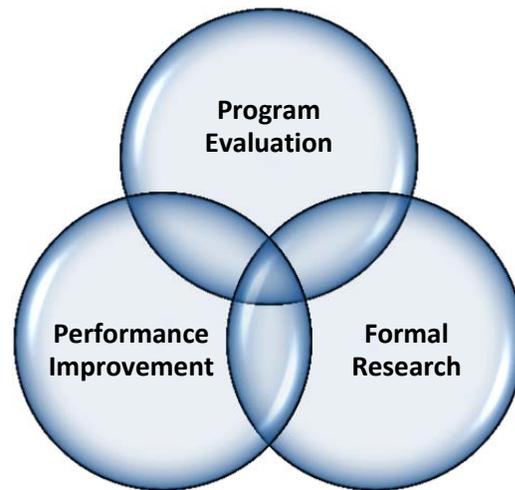
Figure 3. Selecting a Type of Program Evaluation



DISTINCTIONS BETWEEN PROGRAM EVALUATION, PERFORMANCE IMPROVEMENT AND FORMAL RESEARCH

A common question about program evaluation is "How is program evaluation different from the existing performance improvement efforts at our organization?" A related question is "How does program evaluation relate to research?" There are certainly degrees of overlap between program evaluation, performance improvement and formal research on a program's effectiveness. The relationship is depicted in **Figure 4. Relationship Between Program Evaluation, Performance Improvement and Formal Research.** While all of these efforts are similar in that they all can be used to understand and improve the functioning of a program, there are clear differences between them, as described below.

Figure 4. Relationship Between Program Evaluation, Performance Improvement and Formal Research



Program Evaluation – As discussed above, a program evaluation is the process of conducting a systematic study of a program to assess how well it is working. While program evaluation may employ many elements seen in research, it is generally not considered to be “formal” research; in the sense of needing to have full IRB approval (although an IRB exemption may be required).

Performance Improvement – Performance improvements a systematic framework for measuring specific internal processes within an organization. Performance improvement differs from program evaluation in several ways, centering on scope and complexity of the efforts. While program evaluation focuses on a holistic overview of an entire program with the goal of determining if the program is reaching its intended goals, performance improvement efforts are more focused, with an assessment of specific processes within a program. Performance improvement can also be more flexible compared to program evaluation or research, with the ability to add or drop measures as needed; whereas both research and program evaluation efforts require a formalized data collection process for a set period of time. Another difference is that since the aim of performance improvement is to improve the workings of some aspect of the program: the results of performance improvement efforts are typically only shared with the program staff or program owners. Despite the differences, there is clearly an overlap between the two activities in that they both seek to improve a program’s functioning. Program evaluation and performance improvement efforts might also use similar techniques to assess aspects of a program’s functioning. For example, an inpatient ward might track readmissions within seven days as a metric as part of its performance improvement program. If a program evaluation were to be conducted on this ward, it is possible that this same measure would be used as a metric.

Formal Research (FR)—Research on programs that provide services is an essential component in the identification of best practices and evidence-based treatment approaches. Formal

research projects are able to leverage more complex methodologies and statistical tools to answer questions about the effectiveness of a program. Formal research designs allow the effects of extraneous variables to be controlled, which can greatly increase one’s confidence that results are indeed due to the program’s interventions. While formal research offers the greatest level of assurance regarding the validity of the intervention’s effects, conducting formal research is an expensive and complex task, which is outside the capability of many programs. Formal research also requires the approvals and oversight of an Institutional Review Board (IRB), which can add to cost and time required. Research findings are intended for a different audience than that of performance improvement and program evaluation. Formal research seeks to inform the broader scientific community, while program evaluation and performance improvement efforts usually are meant to inform internal decision makers.

REASONS FOR INITIATING PROGRAM EVALUATIONS

Program and portfolio leaders, as well as other stakeholders and decision-makers, may initiate program evaluations for various reasons. Portfolio leaders may request evaluations to better understand the effectiveness of individual programs with the aim of balancing or growing the portfolio. Program leaders may initiate evaluations to highlight successes and to drive self-improvement, while senior Military Health System (MHS) and Veterans Health Administration (VHA) leaders may request evaluations to inform policymaking, budgeting and decision making. Leaders may also initiate evaluations to provide DCoE with documentation of innovations and best practices for dissemination across the MHS. Keep in mind that it is important to plan program evaluations during the initiation stages of new programs.

Portfolios vs. Programs

- A program consists of a group of staff who are engaged in related activities and projects aimed at a particular goal.
- A portfolio encompasses a set of related programs for which an organization provides funding or of which it has oversight.

TEAM COMPOSITION

The size and composition of the evaluation team will vary, depending on size of the program and scope of the evaluation. In many cases, small programs can be evaluated by a single individual who will serve many roles. For larger programs, a team of people may be needed, bringing together people with a variety of skill sets to complete the program evaluation. Based on evaluation complexity, DCoE suggests varying team size and composition as outlined below. Each team ideally includes one team lead, one subject matter expert (SME) and one analyst. As evaluation complexity increases, team size should progressively increase as well. The general responsibilities associated with each of the suggested roles are described in **Figure 5. Evaluation Team Roles and Responsibilities.**

Figure 5. Evaluation Team Roles and Responsibilities

ROLE	RESPONSIBILITIES
Program Lead	<ul style="list-style-type: none"> • Support rationale behind conducting the program evaluation • Maintain active dialogue and participation in the process

ROLE	RESPONSIBILITIES
Team Lead	<ul style="list-style-type: none"> • Manage the team and execute the program evaluation methodology • Apply subject matter expertise and review team insights and analyses • Function as the main communication liaison with the program leaders • Provide feedback on evaluation results to leadership • Direct, review and approve program evaluation outputs • Facilitate discussions during interviews, video and telephone conference calls and team meetings
Subject Matter Expert (s)(SME)	<ul style="list-style-type: none"> • Provide PH, TBI and/or clinical program development and effectiveness expertise and insights • Interpret and analyze program information, clinical publications, reports and other data provided by the program • Study program budget and financial data to conduct cost analysis (if applicable) • Draft findings for outputs • Review draft of the reports for technical accuracy and insight • Support team lead facilitation of interviews and meetings • Assist analyst with documentation of program information obtained from interviews
Analyst	<ul style="list-style-type: none"> • Monitor and report to team lead program evaluation methodology timelines, deadlines and milestones • Confirm scope and desired outcomes for each program evaluation methodology step and template with team lead • Track progress and identify risks for team lead regarding program evaluation methodology steps and templates • Schedule, coordinate and draft agendas for team meetings and conference calls • Support development of methodology outputs using templates; review outputs for accuracy, consistency and clarity • Assure or conduct data collection, collation, data base storage, statistical analysis and provide statement of results. Assist in preparing an analysis of stakeholders and draft initial report of results • Participate in interviews and take notes on information obtained

A PROGRAM EVALUATION FRAMEWORK

This Program Evaluation Guide provides the DCoE framework for planning and implementing a program evaluation through eight sequential steps.

These steps are divided into three distinct phases: Preparation Phase, Execution Phase, and Feedback Phase.

Figure 6. DCoE Program Evaluation Framework



The **Preparation Phase** initiates the evaluation process and consists of four steps. Completing these steps allows program evaluators to gain insight into the program background and current operations, generate buy-in from stakeholders, formulate evaluation questions and develop an evaluation design and data plan relevant to the goals and objectives of the program.

- **Step 1: Review of Program:** This step begins with gathering detailed information about the program, such as the program's background, including its mission, objectives, goals, challenges, and successes. Next, a stakeholder analysis is conducted to understand the individual perspectives and concerns of the various groups that have an interest in the program. Reviewers will also create a logic model of the program detailing the inputs, outputs and anticipated outcomes for the program.
- **Step 2: Develop Evaluation Questions:** This step uses the information obtained in Step 1 to determine what type of evaluation (Process or Outcome) to select, and to prepare a listing of the potential evaluation questions, which will then need to be narrowed to a manageable number. All final evaluation questions will be operationalized using SMART (Specific, Measurable, Achievable, Relevant, Time-bound) criteria, and a specific measure or metric will be selected for each.
- **Step 3: Develop Evaluation Design:** This step begins with a discussion of three types of evaluation designs (descriptive, experimental, and quasi-experimental). The type of study design will be selected during this step, which will guide the formation of the data plan in Step 4. A discussion of internal validity and threats to validity is also provided.
- **Step 4: Develop Data Plan:** This step begins the creation of a data sampling plan, which includes information related to how and when data will be collected. The need for protocols for securing data and the need for standard operating procedures that guide data collection are reviewed. Approvals that may be necessary before beginning the evaluation are discussed, as is a description of setting up a database. Finally, information on training needs and when and how to conduct a pilot test of new procedures is presented. This step concludes the Preparation Phase.

The **Execution Phase** consists of two steps during which the design and plans formulated in Phase I are put into action. These phases involve 1) Gathering Data and 2) Analyzing Data. These steps consist of employing the evaluation design to carry out the evaluation, collecting relevant data, and examining the data gathered to synthesize the evaluation findings for reporting evaluation results.

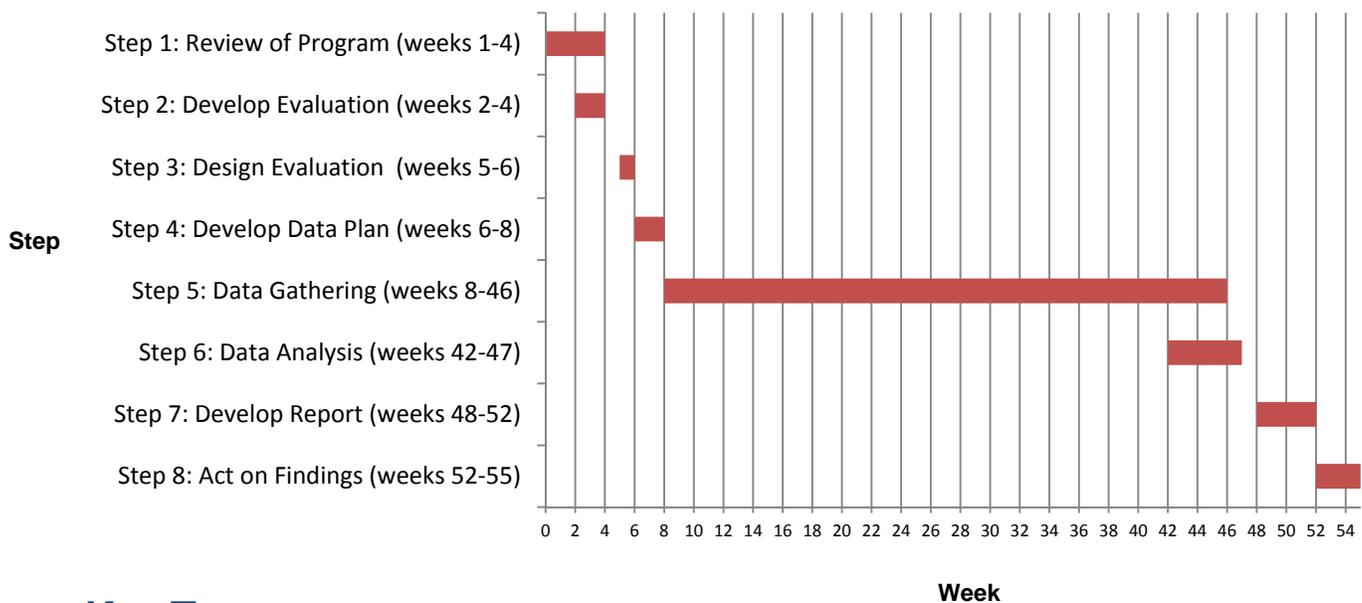
- **Step 5: Gather Data:** The purpose of data gathering is to obtain information to keep on record, make decisions about important issues, and pass information on to others. A data monitoring plan is created to ensure that frequent reviews of the data occur throughout the data collection process to ensure the quality of the data.
- **Step 6: Analyze Data:** This step includes the process of organizing, classifying and interpreting collected data with the goal of uncovering useful information and drawing conclusions to support decision-making. This section provides a description of qualitative and quantitative data, and how data should be prepared before statistical analyses are conducted. Both descriptive and inferential statistics are discussed, as well as information on interpreting the results of the analyses. The completion of data analysis concludes the Execution Phase.

The **Feedback Phase** consists of two steps, 1) Develop Report and 2) Acting on Findings. During this phase, evaluators produce a report of the evaluation findings suitable for sharing with stakeholder groups. The section also involves applying evaluation results to improve program services.

- **Step 7: Report on Findings:** In this step, evaluators are required to determine if the program is meeting its objectives, and synthesize the results of the program evaluation into a formal report.
- **Step 8: Act on Findings:** In this step the information retrieved and analyses performed are acted upon. In this phase actionable decisions about the program are made to improve the program and identify options for moving forward.

The amount of time that each of these steps will require depends on the complexity and type of the evaluation. Process evaluations that have access to data that is already being collected may be completed quickly. Conversely, an outcome evaluation which is collecting data on measures of effectiveness over time may be conducted over a period of years. It is recommended that a timeline for the evaluation be completed which specifies how much time each step is anticipated to take. **Figure 7. Example Program Evaluation Timeline** provides a sample guideline for program which conducted a 55-week evaluation, showing the length of the eight evaluation steps described above.

Figure 7. Example Program Evaluation Timeline



KEY TAKEAWAYS

- Program evaluations are an essential component of delivering quality services to a program's participants and yield numerous benefits.
- It is important to follow a structured approach to program evaluation, such as the DCoE program evaluation process which provides a step by step framework for conducting basic evaluations.

- This Program Evaluation Guide is designed to be an easy-to-use to reference/resource that evaluators can use to perform evaluations of a wide variety of clinical and support programs.

Sources and suggested readings

- Aubel, J. (1999). *Participatory program evaluation manual*. Calverton: Child Survival Technical Support Project and Catholic Relief Services.
- Chinman, M., Imm, P., & Wandersman, A. (2004). *Getting to outcomes: Promoting accountability through methods and tools for planning, implementation, and evaluation*. Arlington: RAND Corporation.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation alternative approaches and practical guidelines (3rd ed.)*. Boston: Pearson Education, Inc.
- Gelmon, S. B., Foucek, A., & Waterbury, A. (2005). *Program Evaluation: Principles and Practices (2nd ed.)*. Portland: Northwest Health Foundation.
- Holden, D. J., & Zimmerman, M. A. (2009). *A Practical Guide to Program Evaluation*. Thousand Oaks, CA: SAGE Publications, Inc.
- Lichiello, P., Turnock, B. J., & Thompson, J. T. (1999). *Guidebook for performance measurement*. University of Washington, Turning Point National Program Office.
- National Minority AIDS Council. (n.d.). *Organizational Effectiveness Series: Program Evaluation*. Retrieved from: <http://www.nmac.org/pdf/program%20evaluation.pdf>
- National Institute on Drug Abuse. (1993). *How good is your drug abuse treatment program?* U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy, Education, and Legislation, Community and Professional Education Branch.
- United Way of America Task Force on Impact. (1996). *Measuring program outcomes: A Practical Approach*. United Way of America.
- W.K. Kellogg Foundation. (1998). *Evaluation Handbook [Item #1203]*. Battle Creek, MI: Collateral Management Company.
- Wolfe, B.L., & Miller, W. R. (n.d.). *Program evaluation: A do-it-yourself manual for substance abuse programs*. Albuquerque, NM: University of New Mexico, Department of Psychology.

Program Evaluation Methodology: Phases and Steps

PHASE 1: PREPARATION



Proper preparation before beginning to conduct the program evaluation is a vital phase. This is when important information concerning the program is gathered. Prior to conducting evaluation activities it is important to develop a clear understanding of how the program is addressing needs and how best to evaluate the program to ensure stakeholder expectations are being met. The Preparation Phase is the first of the three program evaluation methodology phases and consists of four steps:

- Step 1: Review of Program
- Step 2: Develop Evaluation Questions
- Step 3: Develop Evaluation Design
- Step 4: Develop Data Plan

Following these steps in order and applying the precepts covered below will ensure a solid foundation needed to conduct the evaluation, assess the findings, and deliver a meaningful and comprehensive report to the program leaders.

Step 1: Review of Program



The purpose of this first step is to conduct research and interviews in order to familiarize the evaluation team with the basics of the program, such as:

- The program’s background, to include the mission, goals and objectives
- The stakeholders; i.e., those groups of people who participate in, or have an interest in, the program being evaluated
- The overall picture of the program, by creating a logic model that captures the inputs, outputs and outcomes of the program

Review of Program

- Gather information on the program background (charter, mission, goals, objectives)
- Conduct stakeholder analysis
- Create a logic model for the program

Why conduct this step? This is a critical step, especially for reviewers that are external to a program, but should not be glossed over by staff conducting self-evaluations. When conducting a self-evaluation, it may be tempting to assume that all staff has the same picture of the program, but this may be incorrect. How long has it been since the charter, mission, goals, and objectives have been reviewed? If, for example, staffing levels have changed since the program began, or if there are other new and similar programs, this is the time to ensure that everyone on the evaluation team understands the original program framework. Gathering detailed information on paper as to how the program originated, how the program currently operates, who participates in the program, and the perceptions of those participants will set the stage for the remainder of the steps and will focus the evaluation on the areas of most concern to the program leaders.

Template A: Background Review
(double click on image to open)

TEMPLATE A. BACKGROUND REVIEW

This template is intended to guide program leaders through the process of reviewing the program background. This involves conducting a high-level examination of the problem, identifying existing capabilities to address the problem by conducting a leadership discussion of the current program, and comparing it to other similar existing programs. When completing this form, it may be beneficial to use an interview format to gain the best insights into the program.

Questions Used to Review Program Background

Document Desired Result

What is the desired result of this program?

Document Current State

To document the current state, interview program leaders to answer the following questions:

- How and when did the program begin? For example, was the program established as a result of a law, congressional mandate, higher headquarters mandate, or to answer a local area of concern?
- What does leadership perceive is lacking in terms of addressing the identified need or dealing with the problem?
- What are the challenges and successes of this program?
- What similar programs exist which address the area of concern? What lessons learned, if any, can be applied from existing programs?
- What additional services and resources are not currently available to help address the problem?

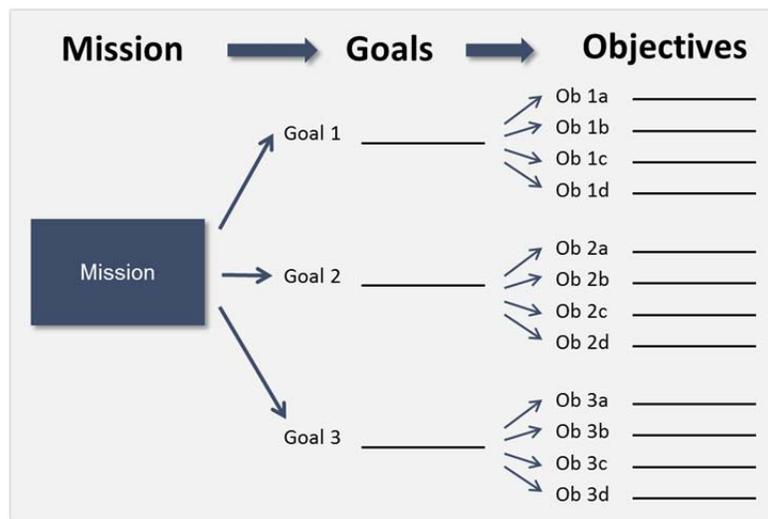
During this step, it may also be helpful to identify any data the program is currently collecting. There may already be sources of information available that will prevent unnecessary work and time collecting data. For instance, in military treatment facilities, there is a wealth of data on patient demographics and throughput (number of appointments, wait times, etc.) available via Composite Health Care System (CHCS) and Armed Forces Health Longitudinal Technology Application (AHLTA). Also, many programs have a comprehensive intake process that includes several standardized measures to assist with diagnosis, which can potentially be used again for measuring change in symptom levels. If the program has undergone a previous program evaluation cycle, it would be advantageous to review the findings and lessons learned. Lastly, if there are any staff members with experience in program evaluation, it would be helpful that their experience be leveraged to help provide expertise and know-how in executing a program evaluation.

Figure 8. Examples of Mission Statements

Program	Mission Statement
Hypothetical Examples	
Traumatic brain injury program	We provide an environment where we work collaboratively to ensure all Wounded Warriors who have experienced traumatic brain injuries are afforded expert state-of-the-art care, compassion, and support throughout the rehabilitation process.
Psychological health program	We are dedicated to the evaluation and care of Wounded Warriors with Psychological Health disorders, ensuring that patients have access to evidence based treatment interventions in a caring environment.
Substance use disorders program	We address the problem of substance abuse among service members through use of intensive inpatient and residential services that stress developing new behaviors that support a drug-free lifestyle.
Suicide prevention program	We address the problem of suicide by providing accurate information to service members and their families about the risk factors and ways to intervene when someone is suicidal, as well as providing support in the form of crisis intervention and referral.
Institutional Examples	
National Intrepid Center of Excellence (NICoE)	As the Military Health System institute dedicated to understanding complex, comorbid traumatic brain injury and psychological health conditions, we deliver comprehensive and holistic care, conduct focused research, and export knowledge to benefit service members, their families and society.
San Antonio Military Medical Center	We are dedicated to the rehabilitation of Wounded Warriors with traumatic brain injury.
Defense Centers of Excellence (DCoE)	Improving the lives of our nation's service members, families and veterans by advancing excellence in psychological health and traumatic brain injury prevention and care.

The mission of the program is the overarching purpose for the program's existence, and the goals and objectives of the program should directly support this mission. While a program will have only one mission, it will typically have several goals. Expanding outward, there may be many objectives to achieve each goal. An illustration of this relationship is seen in **Figure 9. Relationship Between Mission, Goals and Objectives.**

Figure 9. Relationship Between Mission, Goals and Objectives

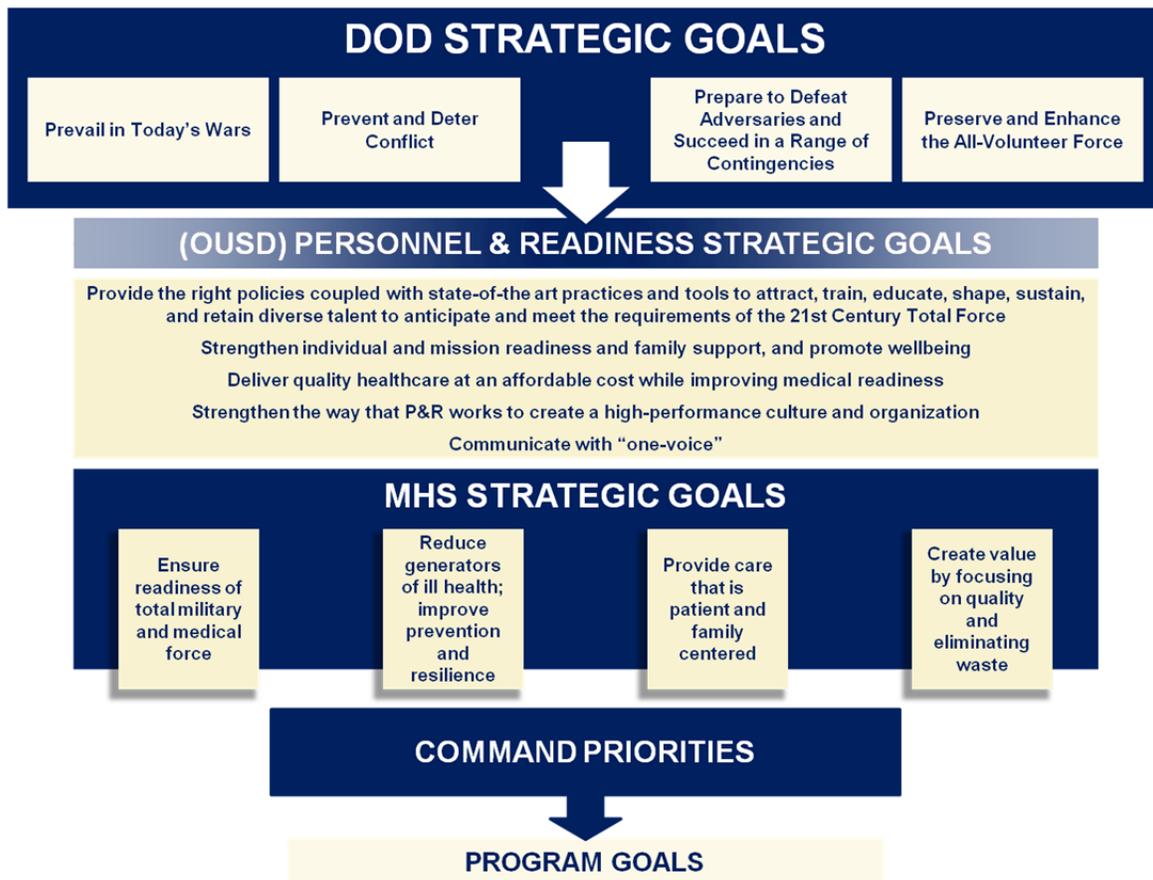


Goals

Once program leaders agree on the mission, the next task is to identify actionable program goals. Goals are statements of aspiration outlining what the program intends to accomplish. They create a framework for determining the specific objectives of a program. Program goals should always be consistent with Defense Department strategic goals as well as the goals, mission and vision of the organization, as depicted in **Figure 10. Developing Program Goals** below. Goals should be broad, overarching statements, whereas objectives break goals down into measurable components.

- Sample Program Goals:**
- Provide an effective and safe treatment program that comprehensively meets the unique needs of active-duty service members with substance dependence disorders.
 - Increase readiness and functioning of active-duty service members with psychological health disorders.
 - Address unique needs of families of active-duty service members with traumatic brain injuries.

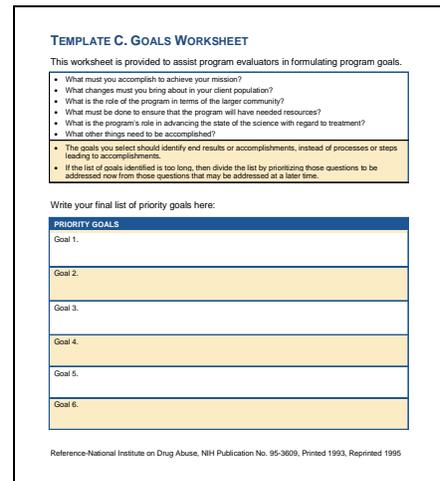
Figure 10. Developing Program Goals



Program leaders should have incorporated various perspectives (e.g., beneficiary or command) as goals were developed. The program may have several goals, and each goal will translate into multiple objectives and effectiveness measures. Program leaders therefore should limit the number of goals examined in order to manage the scope of the evaluation.

What if the findings indicate that the goals are nebulous, do not reflect the mission, or there are no goals? In order to obtain consensus and continue with the evaluation, it will be helpful to work with the program leaders to articulate the goals in a way that will allow the evaluation to focus on the most important issues. **Template C: Goals Template**, will guide program leaders in crafting program goals.

Template C: Goals Template
(double click on image to open)



Objectives

Goals serve to form a bridge between the broad language of a mission statement and the concreteness of program objectives. Once program goals have been established or reviewed, the next step is to define clear, concrete objectives. While it is important for objectives to align with program goals, the command mission and the needs of the audiences must also be considered so that the program will be relevant to those it intends to serve. The purpose of the objectives is to operationally define the overarching goals of a program by breaking them down into smaller, measurable items. Multiple objectives may exist for each program goal. This section provides guidance on how to use the SMART framework to develop objectives that are Specific, Measurable, Achievable, Relevant, and Time-bound.

The SMART framework assists in composing objectives that incorporate measures of program effectiveness based on overarching goals. **Figure 11. The SMART Framework** provides an overview of the SMART framework and additional information about each element of the methodology is provided below.

Figure 11. The SMART Framework

S (SPECIFIC)	M (MEASURABLE)	A (ACHIEVABLE)	R (RELEVANT)	T (TIME-BOUND)
<ul style="list-style-type: none"> Detailed, well-defined Do the objectives specify what the program needs to achieve? 	<ul style="list-style-type: none"> Numeric, observable Can whether or not the program is meeting the objectives be measured? 	<ul style="list-style-type: none"> Actionable, appropriate Can the objectives reasonably be attained given available resources? 	<ul style="list-style-type: none"> Considers population needs Do objectives relate to the program's primary goals? 	<ul style="list-style-type: none"> Defined end point By when do the set objectives need to be achieved?

Specific. Program objectives must be specific and concrete, not general or abstract. Developing specific objectives will assist in meeting the other criteria in the SMART framework.

Measurable. A key component of objective development is to identify which aspects of the program goals can be quantifiably measured. This will allow for the collection of data that will statistically measure the effectiveness of a program.

Achievable. An achievable objective is actionable and appropriate. It should also be possible to meet this objective given the available resources.

Relevant. Relevance to program goals and to audience needs must also be considered. This aspect of the SMART framework recognizes the reality-based nature of the objective. The task here is to select objectives that most directly and clearly indicate that the program is meeting its goals.

Time-bound. An objective should have a defined end point and/or a specified length of time.

Well-constructed objectives provide program leaders with a measurable way to determine if the program is meeting its stated goals. Because objectives are established to support achievement of program goals, the language in the objectives must be concrete, action-oriented and characterized by a desired outcome or end state. **Template D: Objectives Development** can be used to determine whether or not a program’s objectives are consistent with the SMART criteria.

Template D: Objectives Development
(double click on image to open)

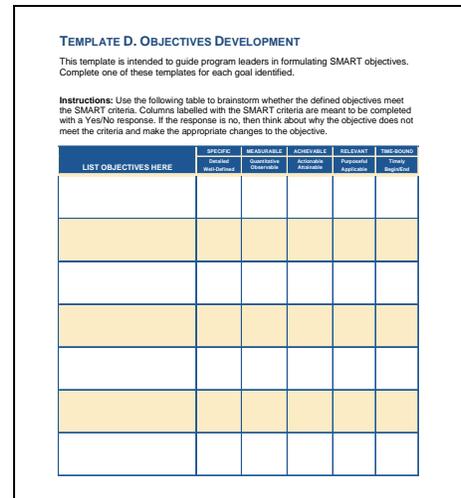


Figure 12. Example Objectives provides an example of both an effective and an ineffective objective for a sample program.

Figure 12. Example Objectives

OBJECTIVE	SMART?	EXPLANATION
Program participants will show a decrease in substance abuse	No	This objective might be achievable, but it is not specific and is not relevant to population needs. Additionally, it does not indicate a time period in which symptoms will decrease.
Participants will maintain abstinence from alcohol and other substances, as measured by self-report and weekly drug testing, during the 28-day program.	Yes	This objective describes a change within a specified period of time. This objective is also achievable and relevant. It also details how the outcome will be measured.

Stakeholder Analysis

Once the background of the program being evaluated has been established, the next task is to identify and interview the program stakeholders. A stakeholder is any person or group who has an interest in the evaluation or the evaluation results. These groups include staff members, patients, agencies the program interacts with or reports to, or community-based advocacy groups. Despite having already gathered information from the program leaders during the previous task, it is important to understand that there will likely be more than one group, with different perspectives of the program.

Template E: Stakeholder Analysis provides the evaluation team an opportunity to identify those stakeholder groups and set the tone for a collaborative relationship with these groups. A strong relationship with stakeholders facilitates smooth execution of the program evaluation methodology and focuses the evaluation and its results so that they are meaningful to the evaluation team, the program leaders, and the stakeholders. Engaging with as many stakeholder groups as is possible and feasible will help to ensure that the evaluation design addresses their concerns. Therefore, a strategy of how best to engage each stakeholder should be discussed and documented in the template.

Completing the Stakeholder Analysis step can be cumbersome, but it will provide important insight into how the program is functioning through the eyes of staff, patients, and any other group that interacts with the program. For example, including staff members in the stakeholder analysis will give the evaluation team insight into areas of concern of which the leaders may not be aware and it will help to gain buy-in from key staff members whom may provide support during the Gather Data step.

If the evaluation team consists of individuals external to the program and its staff, regular communication with the primary program leaders helps the team to maintain transparency and clearly communicate expectations. Building the stakeholder engagement early in the evaluation methodology provides the team with opportunities to understand the individual stakeholder group's interest in the program and solicit their input.

Depending on how the evaluation is initiated, stakeholders may not fully understand the evaluation process. A **Stakeholder Engagement Presentation [Template F]** can be used as the first communication between the team lead and each of the evaluation stakeholder groups. Even in cases of self-evaluation, the presentation provides an opportunity to both inform stakeholders about the evaluation and to gain support for the evaluation. The team lead will use the presentation to introduce the rationale for conducting the evaluation, describe the program evaluation methodology and solicit stakeholder input respective to their interests in the program. The Stakeholder Engagement Presentation also provides an opportunity for the team to ask background questions and request program documents to prepare for the remaining steps in the Preparation Phase.

The presentation includes the following information:

- Rationale for conducting program evaluation
- Types of evaluation approaches
- Suggested timeline(s) and key dates for the evaluation
- Expected dissemination strategy for evaluation results
- Details in the Review of Program template
- Request for input on stakeholder interest in the program and program evaluation

Template E: Stakeholder Analysis
(double click on image to open)

TEMPLATE E. STAKEHOLDER ANALYSIS
To complete the stakeholder analysis, interview each identified stakeholder group, completing the fields in the form below.

Stakeholder name	Role or relation to program	Use of program materials or services	Areas of concern for this stakeholder	Follow-up meetings planned or requested

Template F: Stakeholder Engagement Presentation
(double click to open)

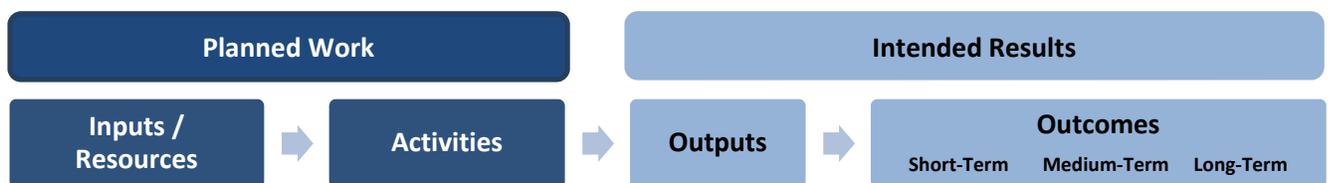
[Organization Conducting Evaluation]
Program Evaluation Stakeholder Engagement Presentation
[Program Name (Acronym)]
DD MM YYYY

The Stakeholder Engagement Presentation also enables the team to establish a positive first impression and discuss questions or concerns that the primary stakeholders may have about the program evaluation. Typically, stakeholders would like to know who will receive the final evaluation report and how information contained in the report may impact the program’s future. Depending on the evaluation’s purpose, the answers to stakeholder questions vary; however, to facilitate a smooth evaluation process, the team lead should use this opportunity to emphasize the transparent and collaborative nature of the evaluation methodology.

Create a Logic Model for the Program

The final task in the Review of Program step is to create a logic model of the program. A logic model graphically details how a program is structured and how it intends to achieve its expected results. Representing a linear sequence, a logic model includes four key components: inputs/resources; activities (planned work); outputs; and outcomes (intended results). At the simplest level, a logic model displays the resources a program is using in its outputs that produce a set of outcomes. The more information gathered during this task, the more prepared the team will be to conduct the actual evaluation. **Figure 13. Logic Model Components** depicts the domains included in a logic model.

Figure 13. Logic Model Components



A program’s *inputs/resources* reflect the resources available to, and utilized by, a program. In other words, inputs are put *in* a program. Resources can be financial, physical or human and can include funding, staff, volunteers, and equipment.

Activities are what the program does with those inputs/resources.

Outputs capture the products (in types, levels, or targets) of program’s activities. They may be thought of as units of service.

Outcomes reflect the impact of the program and can vary in scope from short and medium-term (such as knowledge, skills and individual behavior) to long-term (such as social and environmental impacts).

A logic model provides a visual and conceptual link, or bridge, between Step 1 and Step 2, assisting the evaluation team to determine which aspects to focus on when developing the evaluation questions. The evaluation team should complete a logic model for the program, working with the program staff and stakeholders to gain the most complete picture of the program possible.

This Guide has four separate templates for helping staff brainstorm how to generate the specific components for each of the logic model domains (Inputs, Activities, Outputs and Outcomes).

- **Program Inputs Brainstorming Sheet [Template G]**
- **Program Activities Brainstorming Sheet [Template H]**
- **Program Outputs Brainstorming Sheet [Template I]**

- **Program Outcomes Brainstorming Sheet [Template J]**

After the evaluation team has completed the worksheets for inputs, activities, outputs, and outcomes, this information can be distilled into a logic model that summarizes the program from a holistic viewpoint. [See **Template K: Logic Model Template**]

Template G: Program Inputs Brainstorming Sheet
(double click on image to open)

TEMPLATE G. PROGRAM INPUTS BRAINSTORMING SHEET
Use this sheet to generate information on the program's resources, which can include staff, space available, levels of funding, and equipment.

STAFF: Full-Time Part-Time Equipment Type (MD, RN, PT, OT, etc.)	
SPACE: Dedicated Shared	
BUDGET: Salaries Training Travel Marketing	
SUPPLIES / EQUIPMENT: Medical Office	
OTHER:	

Template H: Program Activities Brainstorming Sheet
(double click on image to open)

TEMPLATE H. PROGRAM ACTIVITIES BRAINSTORMING SHEET
In generating a list of the program outputs, consider the all of activities that staff engage in on a day-to-day basis. The table below lists several example activities, and provides a space for you to document your own programs outputs.

PROGRAM ACTIVITIES		
Treatment Planning Medication Management Diagnostic Testing Patient Follow-Up Calls Field Trips / Outings Patient Education	Individual Treatment Sessions Group Therapy Sessions Workshops Classes Documentation / Charting Interviews	Command Liaison Distributing Materials Developing Public Service Announcements Website Hosting
1)		
2)		
3)		
4)		
5)		
6)		
7)		
8)		
9)		
10)		
11)		
12)		
13)		
14)		
15)		
16)		
17)		

Template I: Program Outputs Brainstorming Sheet
(double click on image to open)

TEMPLATE I. PROGRAM OUTPUTS BRAINSTORMING SHEET
The outputs of a program should tie back to the Program's Activities; they provide an operational definition of the activity which can be quantified. Some examples are provided below:

PROGRAM ACTIVITIES	PROGRAM OUTPUTS (MEASURES OF ACTIVITIES) (if available)
Individual Treatment Sessions	# of individual treatment sessions per month (participant volume) average wait time until next available therapy appointment (capacity) number of unfilled therapy appointments (excess capacity)
Classes	# of classes provided per month # of attendees per group and total

Template J: Program Outcomes Brainstorming Sheet
(double click on image to open)

TEMPLATE J. PROGRAM OUTCOMES BRAINSTORMING SHEET
In generating a list of the program outcomes, consider outcomes that span the immediate (short-term) to distant (long term) time frame. Examples of potential outcomes within each timeframe for each are shown below.

SHORT-TERM	PROGRAM OUTCOMES	
	MID-TERM	LONG-TERM
Knowledge Levels Skill Levels Attitudes	Decrease in Symptoms Increased Coping Skills Decrease in Risk Factors	Decreased Disability Absence of Disorder Improved Quality of Life
1.	1.	1.
2.	2.	2.
3.	3.	3.
4.	4.	4.
5.	5.	5.

Template K: Logic Model Template
(double click on image to open)

TEMPLATE K. LOGIC MODEL TEMPLATE					
← Process Evaluations →			← Outcome Evaluations →		
INPUTS	ACTIVITIES	OUTPUTS	OUTCOMES		
			SHORT-TERM	MID-TERM	LONG-TERM

A completed evaluation logic model will include the range of inputs, outputs, and outcomes reflective of the specific program of interest. **Figure 14. Example Logic Model for a Specialty Clinic for Combat-Related PTSD, Figure 15. Example Logic Model for a Specialty TBI Clinic, and Figure 16. Example Logic Model for a Substance Abuse Clinic** present examples of completed evaluation logic models. Typically, an outcome evaluation will focus on the outcomes or goals of the program, while a process evaluation will focus on inputs and outputs. Types of evaluation questions are discussed in **Step 2: Develop Evaluation Questions**.

Conclusion

Review of Program is the first step in Program evaluation and involves establishing an understanding of the program, engaging and building support with stakeholders, as well as identifying existing resources, program activities, participants, and expected outcomes of the program.

The Review of Program step is vital to performing a successful evaluation. This step may well be the most time consuming step in the evaluation process; however, investing the time up front to learn about these items in preparation for conducting the evaluation, will ultimately save time downstream and allow for meaningful evaluation.

Key Takeaways

- First Step: Gather detailed information about the program before conducting the evaluation
- Program Background: program leaders can provide an overview of program, challenges and successes
- Stakeholder Analysis: understand there may be more than one group, with different perspectives
- Logic Models: identify the inputs/resources that the program requires, the activities it plans to achieve, the outputs of the program, and the anticipated outcomes.

Sources and suggested readings

Lichiello, P., Turnock, B. J., & Thompson, J. T. (1999). *Guidebook for performance measurement*. University of Washington, Turning Point National Program Office.

National Institute on Drug Abuse. (1993). *How good is your drug abuse treatment program?: Overview and case study*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy, Education, and Legislation, Community and Professional Education Branch.

United States Government Accountability Office. (2012). *Designing evaluations* (2012 Revision). 12-208G.

United Way of America Task Force on Impact. (1996). *Measuring program outcomes: A practical approach*. United Way of America.

W.K. Kellogg Foundation. (1998). *Evaluation Handbook [Item #1203]*. Battle Creek, MI: Collateral Management Company.

W.K. Kellogg Foundation. (2004). *Logic model development guide: Using logic models to bring together planning, evaluation, and action*. Battle Creek, MI: W.K. Kellogg Foundation.

Figure 14. Example Logic Model for a Specialty Clinic for Combat-Related PTSD

INPUTS (resources available)	ACTIVITIES (what program does)	OUTPUTS (measures of activities)	OUTCOMES		
			SHORT-TERM	MID-TERM	LONG-TERM
Staff: <ul style="list-style-type: none"> • 4 Clinical Psychologists • 4 Licensed Clinical Social workers • 2.5 Psychiatrists • 4 Psychiatric Technicians • 2 Booking clerks Space: <ul style="list-style-type: none"> • 2 Group rooms (shared with another program) • 15 office spaces Funding: <ul style="list-style-type: none"> • 5.6 million per year Equipment/Supplies: <ul style="list-style-type: none"> • General office supplies • Biofeedback setup (2) 	• Intake assessments	# of intakes/month; Wait times for intake (by provider type)	<ul style="list-style-type: none"> • Increased knowledge about PTSD • Increased knowledge about effective treatments • Acquisition of coping skills / knowledge of how to apply therapeutic exercises • Satisfaction with overall program 	<ul style="list-style-type: none"> • Decreased levels of symptoms of disorders (PTSD, MDD, SUD). • Increased ability cope with day-to-day stressors • Decreased levels of risky behaviors (driving while intoxicated) 	<ul style="list-style-type: none"> • Decreased disability due to deployment related illnesses • Absence of disorder (patients no longer meet criteria) • Decreased risk for administrative separation for problematic behaviors (alcohol use, anger outbursts, etc.) • Improved Quality of Life
	• Medication management	# of appointments/month, wait time till next available appointment			
	• Treatment planning	percentage of patients with completed treatment plans			
	• Individual therapy	# of sessions for each type; wait time for next available appointment by type			
	• Marital / family counseling				
	• Biofeedback training				
	• Group therapy sessions	# of sessions; # of attendees per group or class; wait times for group or class			
	• Psychoeducational Classes				
	• Multidisciplinary treatment team meetings	# of meetings per month, number of patients covered/meeting			
• Writing medical boards	# medical boards (LIMDU/MEB)				
• Charting in medical records	percentage of notes completed within chart review standards				
• Command liaison	# contacts with commands				

Figure 15. Example Logic Model for a Specialty TBI Clinic

INPUTS (resources available)	ACTIVITIES (what program does)	OUTPUTS (measures of activities)	OUTCOMES		
			SHORT-TERM	MID-TERM	LONG-TERM
Staff: <ul style="list-style-type: none"> • 1 Program director (Neurologist) • 0.5 Neuropsychologist • 1 Neuropsychiatric technician • 2 Occupational therapists • 1 Certified OT assistant. • 1 Speech language pathologist • 1 Recreational Therapist • 1 Booking clerk Space: <ul style="list-style-type: none"> • 1 Group room • 6 office spaces Funding: <ul style="list-style-type: none"> • 2.4 million per year Equipment/Supplies: <ul style="list-style-type: none"> • General office supplies • OT / SLP supplies 	• Intake assessments	# of intakes/month; Wait times for intake (by provider type)	<ul style="list-style-type: none"> • Increased knowledge about mTBI • Increased knowledge about effective treatments • Acquisition of coping skills / knowledge of how to apply therapeutic exercises • Satisfaction with overall program 	<ul style="list-style-type: none"> • Decreased levels of symptoms of mTBI • Increased ability cope with day-to-day stress • Increased functioning at work and at home 	<ul style="list-style-type: none"> • Decreased disability due to mTBI • Absence of disorder (patients no longer meet criteria) • Decreased risk for administrative separation for problematic behaviors (oversleeping, anger outbursts, etc.) • Improved Quality of Life
	• Medication management	# of appointments/month, wait time till next available appointment			
	• Treatment planning	percentage of patients with completed treatment plans			
	• Cognitive rehabilitation • Occupational therapy • Speech therapy • Vestibular rehabilitation	# of sessions for each type; wait time for next available appointment by type			
	• Diagnostic testing	# of CT scans/MRIs per month			
	• Neuropsychological testing	# of screens & full batteries per month			
	• Multidisciplinary treatment team meetings	# of meetings per month, number of patients covered/meeting			
	• Writing medical boards	# medical boards (LIMDU/MEB)			
	• Patient outings	# of outings, with number in attendance			
	• Charting in medical records	percentage of notes completed within chart review standards			

Figure 16. Example Logic Model for a Substance Abuse Clinic

INPUTS (resources available)	ACTIVITIES (what program does)	OUTPUTS (measures of activities)	OUTCOMES		
			SHORT-TERM	MID-TERM	LONG-TERM
Staff: <ul style="list-style-type: none"> • 1 Psychiatrist-addictions certified • 5 Licensed Clinical Social workers • 3 Interns (2 social work, 1 psychology) • 1 Marital therapist • 0.5 Chaplain • 1 Booking clerk Space: <ul style="list-style-type: none"> • 3 Group rooms • 8 office spaces Funding: <ul style="list-style-type: none"> • 3.1 million per year Equipment/Supplies: <ul style="list-style-type: none"> • General office supplies 	<ul style="list-style-type: none"> • Intake assessments 	# of intakes/month; Wait times for intake (by provider type)	<ul style="list-style-type: none"> • Increased knowledge about SUD • Increased knowledge about effective treatments • Acquisition of coping skills / knowledge of how to apply therapeutic exercises • Satisfaction with overall program 	<ul style="list-style-type: none"> • Decreased levels of symptoms of SUD (lowered cravings, decreased consumption). • Increased ability cope with day-to-day stressors • Improved interactions with family • Decreased levels of risky behaviors (driving while intoxicated) 	<ul style="list-style-type: none"> • Decreased disability due to SUD • Absence of disorder (patients no longer meet criteria) • Decreased risk for administrative separation for problematic behaviors (underage drinking, driving while intoxicated) • Improved Quality of Life
	<ul style="list-style-type: none"> • Treatment planning 	percentage of patients with completed treatment plans			
	<ul style="list-style-type: none"> • Individual therapy • Marital / family counseling 	# of sessions for each type; wait time for next available appointment by type			
	<ul style="list-style-type: none"> • Group therapy sessions • Psychoeducational Classes 	# of sessions; # of attendees per group or class; wait times for group or class			
	<ul style="list-style-type: none"> • Field trips/planned outings 	# of outings; # of attendees per outing			
	<ul style="list-style-type: none"> • Multidisciplinary treatment team meetings 	# of meetings per month, number of patients covered/meeting			
	<ul style="list-style-type: none"> • Writing medical boards 	# medical boards (LIMDU/MEB)			
	<ul style="list-style-type: none"> • Charting in medical records 	percentage of notes completed within chart review standards			
<ul style="list-style-type: none"> • Command liaison 	# contacts with commands				

Step 2: Develop Evaluation Questions



The next step in the Preparation Phase is to **Develop Evaluation Questions** which will aid the program evaluation team in determining the evaluation design and data plan that will best answer the queries. It is best to limit the number of questions in order to tailor the evaluation so that it is manageable and will yield useful information on which to base program decisions. It is important to follow the program evaluation steps in order as described in this guide. Like building blocks, the results of each step will determine the inputs for the steps to follow. **Step 1: Review of Program** (for example), will be used to write the Evaluation Questions during Step 2, which will, in turn, lead to the Evaluation Design (Step 3) and the development of the Data Plan (Step 4).

Recall that during Step 1, the program leaders and the evaluation team reviewed the program’s mission, goals and objectives. As discussed earlier, it is crucial to have consensus on those, and it may have been necessary to guide the program leaders through an exercise to rewrite any that were not agreed to by the group. By the end of Step 1, the evaluation team should have a list of approved SMART program objectives. This is where the process of developing evaluation questions begins. Evaluation questions will allow the team to organize and determine the outcomes to be evaluated and to determine the measures and metrics that will be used to evaluate the program.

Develop Evaluation Questions

- Review information from Step 1, Review of Program
- Determine evaluation type
- Generate evaluation questions
- Operationalize all evaluation questions to SMART criteria
- Select measure or metric for each question

Understanding how and what is important to evaluate can only occur once there is a clear *evaluation goal*, based on sound *program goals and objectives*.

Evaluation questions might come from external sources, such as directly from headquarters or federal agencies. In the absence of, or in addition to those, additional questions will need to be developed. The sections to follow describe factors to consider when assessing prescribed questions or when developing questions.

Comprehension of Program Review

A firm understanding of the mission, goals and SMART objectives of a program, stakeholder engagement, and current resources is critical, as this information is the very foundation for developing sound evaluation questions. Once a firm appreciation of the goals and objectives is achieved, a review of the stakeholder analysis and logic model provides important information needed to shape the evaluation questions. Understanding who requests the evaluation, the intended purpose of the information, and the secondary objectives that may exist serves to guide the focus of the evaluation. Likewise, the list of program resources (staffing, equipment, hours of operation, etc.) provides additional information that can help determine what type of evaluation will be the most impactful. Ensuring a strong understanding for the foundational components of a program is an activity that cannot be over-emphasized. Much like a foundation for a home, this supports each subsequent step in developing evaluation questions.

Determination of Evaluation Type and Focus

Evaluation questions, by nature, can explore any aspect of a program and vary tremendously in both type and focus. Questions corresponding to the three types of evaluations (process, outcome, and impact/cost), discussed in this Guide are described below. The type of evaluation is a key factor in developing evaluation questions. Program evaluations, and therefore evaluation questions, may focus on some or all components of a program, from program inputs to the most long-term intended results. There is no single optimal evaluation focus for a program, as it will differ for each situation – a single program could experience very different evaluation foci over time. The appropriate type and focus for a specific evaluation will depend on the goal of that evaluation. If the team members are not experienced in conducting program evaluations or if staff and resources are limited, evaluations should be kept simple and limited to specific components or populations of the program. Newer programs will likely focus on process evaluations, as they have not likely had time to develop outcome data that can be meaningfully interpreted (see **Figure 17. Scope of Analysis for Process vs. Outcome Evaluations**). Such evaluations will center on the program's inputs and outputs. More mature programs which have outcome data available will likely want to perform an outcome evaluation, which considers inputs, outputs as well as outcomes.

Figure 17. Scope of Analysis for Process vs. Outcome Evaluations



The three types of evaluation questions with examples of each type are displayed in **Figure 18. Sample Evaluation Questions by Evaluation Type.**

Figure 18. Sample Evaluation Questions by Evaluation Type

Evaluation Type	Sample Evaluation Questions
Process Evaluation	<ul style="list-style-type: none"> • How similar are the participants to those anticipated when the program was designed (e.g., age, gender, severity of need)? • Are services being delivered as planned? • How many service hours did each participant receive? • Is the program being implemented as scheduled? • What was the level of participant satisfaction with program services?
Outcome Evaluation	<ul style="list-style-type: none"> • To what extent did the program activities achieve the desired outcomes? • Are there any unexpected effects seen from the program activities? • Were there any unintended (negative) outcomes? • What should be improved or changed in the program? • Did the program impact vary by sub-population?
Impact/Cost Evaluation	<ul style="list-style-type: none"> • What outcomes are attributable to the program as opposed to other internal and external influences? • Which components of the program are responsible for specific outcomes? • Does the benefit of the program to its participants warrant its costs?

Process evaluation questions measure the degree to which a program is performing as it was intended. In essence, process evaluations document program implementation and effort. Such questions can focus on a range of topics, including a program’s conformity to statutory and regulatory requirements, the types and quantities of services delivered, and the beneficiaries of the services. In other words, process evaluation questions focus on the inputs and outputs of a program. The information needed to answer process evaluation questions are typically easier to collect than outcome and cost-benefit as they can often be found in participant or program records. Overall, answers to process evaluation questions are useful in identifying *how* program impact and outcomes were achieved, and can be useful for program replication.

Outcome evaluation questions evaluate *if* the program’s intended outcomes were achieved and can include suggestions on quality and productivity improvements. Fundamentally, outcome evaluations assess the effectiveness of a program to produce change and link outputs to outcomes. This linkage is important as it allows for a reasonable conclusion on the program’s actual success. Outcome evaluation questions are centered on what happened to the program participants and the unique impact the program had on participants’ outcomes. Answers to outcome evaluation questions are important when investigations of a program’s ability to meet a funder’s objectives are required, or for a program that is using an innovative model that has not yet been demonstrated as effective.

Impact Evaluation/Cost Analysis questions assess program benefits, outputs or outcomes and compare them with both the external and internal cost of producing them. The process of developing cost-benefit evaluation or cost-effectiveness evaluation questions can introduce a range of technical challenges and exceeds the scope of this guide. It is important to note that a meaningful cost analysis first requires a successful outcome evaluation.

Selection of Evaluation Questions

One of the most difficult tasks in performing a program evaluation is to narrow the scope of the evaluation to be manageable from a cost and time standpoint, while still yielding sound data

about the program. Potential evaluation questions can originate from the review of the program’s background, mission, goals and objectives. A key source of potential questions for a program evaluation is the program objectives from Step 1. Additional questions can come from stakeholder engagement and the logic model. The evaluator may also be directed to answer specific evaluation questions by the Chain of Command, program staff, or other entities.

Program evaluators should keep in mind the practical realities that go with collecting data, and balance the desire to collect large amounts of data against the costs of having a large amount of data in terms of staff time and later analysis. As a general rule of thumb, only data that is needed and can be analyzed should be collected. The final number of questions that an evaluation team settles on will depend upon a number of factors, including the level of resources for data collection and analysis, the complexity of the questions, etc.

Like program objectives, evaluation questions must also be operationalized as SMART questions. **Figure 19. Sample Evaluation Questions** provides examples of evaluation questions for a sample program offering alcohol treatment to service members. This figure explores how to operationalize SMART evaluation questions in order to provide the greatest degree of information and meaning.

Figure 19. Sample Evaluation Questions

Evaluation Questions	SMART?	Explanation
What is the percentage of service members screened for alcohol use and referred for counseling?	No	This evaluation question is not specific and could be measured in multiple ways, each with different interpretations.
What is the percentage of service members screened with AUDIT-C and referred for brief alcohol counseling?	No	This evaluation question requires a specific screening measure and type of counseling. However, it does not define criteria or a time frame for referral.
What is the percentage of service members (not seen in an alcohol treatment program in past 90 days) screened with AUDIT-C and achieving score of five or greater AND received brief alcohol counseling (feedback linking drinking alcohol to health and advice to abstain or drink within recommended levels)? Counseling must occur within 14 days of the positive screen.	Yes	This evaluation question provides operational definitions and scores for including or excluding veterans. It also provides a precise window of time for the referral in order to ensure that the referral is linked to the screening behavior.

Choosing Data Collection Tools and Measures

Well-developed evaluation questions will also specify how the questions will be answered, noting the tool or measure that will be used to assess the question. In the example in **Figure 19. Sample Evaluation Questions** (above), the evaluation question required the selection of a specific measure for alcohol screening; the AUDIT-C.

In order to select the most appropriate data collection tool or instrument, decision-makers should consider the following factors:

- Cost
- Time to administer
- Specialized training needs
- Reliability and validity

- Requirements/guidance from higher headquarters or governing agencies

There are a variety of data collection options which program evaluators can utilize to gather information about the program, such as self-report measures, population surveys, focus groups, etc. Several types of these tools are listed in **Figure 20. Common Data Collection Tools** (below) with examples of when each might be most useful and some of the advantages and disadvantages of each. There is no one tool that is best in all cases and the choice of what data collection tool(s) to use is dependent on the type of evaluation, type of questions and the evaluation staff’s level of familiarity with the assessment instrument. Strive to use tools that have been tested and proven to be reliable and valid (these concepts will be explained in Step 5).

Figure 20. Common Data Collection Tools

TOOL	WHEN TO USE	ADVANTAGES	DISADVANTAGES
Self-Report Measures	<ul style="list-style-type: none"> • To obtain information directly from respondents • To allow respondents to provide feedback independently 	<ul style="list-style-type: none"> • Obtain participant’s own perspective and account of events 	<ul style="list-style-type: none"> • Subjectivity gives rise to limitations in the validity of data • Risk of underreporting or over-reporting data
Surveys/ Questionnaires	<ul style="list-style-type: none"> • To obtain data from a defined sample population 	<ul style="list-style-type: none"> • Availability of standardized instruments • Allows for anonymous responses • Retrieve data from large groups at one time • Retrieve data quickly • Ease of data analysis due to standardization of responses • Low-cost 	<ul style="list-style-type: none"> • Sample may not be representative of target population • Risk of low response rate • Depending on question wording, may result in biased responses • Depending on question types, may result in insufficient data • Must be written to accommodate all educational levels
Interviews	<ul style="list-style-type: none"> • To obtain information about individual participant’s experiences in detail 	<ul style="list-style-type: none"> • Retrieve detailed information • As compared to surveys, may result in better response rate • Interviewer has flexibility of asking a variety of question types • Allows for observation of gestures, facial expressions, pauses. 	<ul style="list-style-type: none"> • Time-consuming • Requires experienced facilitator • Less anonymity • Qualitative data more difficult to aggregate and analyze • Increased cost of administration

TOOL	WHEN TO USE	ADVANTAGES	DISADVANTAGES
Archival Data	<ul style="list-style-type: none"> To obtain historical program data already available 	<ul style="list-style-type: none"> Data is readily available Program is uninterrupted Can tap into a wide array of data sources, such as medical records or insurance claims data Decreases additional data collection demands on staff and personnel Reduced risk of bias Low cost 	<ul style="list-style-type: none"> Limited by what exists and is available May not find relevant data required Data may be incomplete or have errors Must define data needed
Focus Groups	<ul style="list-style-type: none"> To obtain detailed data about the experiences or perceptions of a defined group 	<ul style="list-style-type: none"> Obtain data from a group at one time Obtain detailed information Group interaction may yield more information As compared to interviews, less costly May be conducted in a short time frame 	<ul style="list-style-type: none"> Requires experienced facilitator May only ask a limited number of questions Group setting may influence or restrict responses Less anonymity Qualitative data more difficult to aggregate and analyze
Case Studies	<ul style="list-style-type: none"> To provide a thorough and comprehensive analysis of a person or event, usually one that is atypical 	<ul style="list-style-type: none"> Useful to provide account of a rare occurrence or event Allows for cross-comparisons Provides an analysis of outliers 	<ul style="list-style-type: none"> Represents only one data point Unable to generalize results to study population Subjective
Program Reports	<ul style="list-style-type: none"> To obtain data already being generated by the unit, department or facility 	<ul style="list-style-type: none"> Data already compiled into weekly, monthly, quarterly or annual reports Data is readily available Program is uninterrupted Less costly 	<ul style="list-style-type: none"> Limited by what exists and is available May not be the relevant data required

After each evaluation question has been framed in operational terms using the SMART criteria, and the evaluation team has determined the type of data collection tool to use in answering the question, the next step will be to select a specific measure. Depending on the type of evaluation being conducted (i.e. process or outcome), there are different types of measures to consider. Output and outcome measures are discussed below, with several examples.

Output Measures for Process Evaluations

Program outputs can be thought of as the operationalization of a program’s activities. For example, if one activity a program lists is providing classes on reintegration for service members, the output for this activity would be the number of classes delivered per month and the attendance (total and per session).

Outputs should be classified using a defined, measurable event. **Figure 21. Sample Output Measures for Process Evaluations** provides example program activities and potential output measures for process-based evaluations.

Figure 21. Sample Output Measures for Process Evaluations

Activities	Potential Output Measures
New participants/patients	<ul style="list-style-type: none"> Number of new intake appointments Wait time for next available intake appointment/by provider type Percentage unfilled intakes (excess capacity)
Caseload	<ul style="list-style-type: none"> Number of patients per provider Number of unfilled follow-up appointment slots
Diagnostic procedures	<ul style="list-style-type: none"> Number of computed tomography (CT) or magnetic resonance imaging (MRI) scans per month Number of swallow tests per month Number of breathalyzer/urinalysis tests per month
Workshops	<ul style="list-style-type: none"> Number of workshop participants Number of workshops per month/year
Online Trainings	<ul style="list-style-type: none"> Number of website views Percentage of participants who complete online training module
Patient demographics	<ul style="list-style-type: none"> Percentage of patients by patient category Percentage of patients by rank / service / gender
Patient characteristics	<ul style="list-style-type: none"> Frequency of types of diagnosis Number of patients in limited duty (LIMDU) status

Outcome Measures for Outcome Evaluations

Outcomes are the results yielded by the program. Outcomes measures are used to ascertain program impact. Outcomes measures can focus on a variety of program results, such as: level of patient functioning or clinical outcomes. For example, evaluators may want to know if patients with PTSD experience a decrease in symptom levels after completing a year-long outpatient PTSD counseling program. This data may be collected from a variety of sources, such as patient surveys, or standardized provider-administered assessments. **Figure 22. Sample Outcome Measures for Outcome Evaluations** provides example outputs and potential measures for outcome-based evaluations.

Figure 22. Sample Outcome Measures for Outcome Evaluations

OUTCOMES	Potential Measures
Decrease in symptoms of disorder	<ul style="list-style-type: none"> Scores on a standardized self-report measure for the disorder in question (e.g., PCL (PTSD check list)) Clinician ratings of improvement in symptoms

OUTCOMES	Potential Measures
Decrease in levels of disability	<ul style="list-style-type: none"> • Number of patients referred to Physical Evaluation Board • Number of patients in limited duty status • Return to duty percentage
Increased coping skills	<ul style="list-style-type: none"> • Scores on self-report measures for coping • Clinician ratings of improvements in coping
Decreased levels of risky behaviors	<ul style="list-style-type: none"> • Amount of alcohol intake • Amount of illicit drug use • Amount of reckless driving
Increased quality of life	<ul style="list-style-type: none"> • Scores on a self-report measure for quality of life
Increase in functioning	<ul style="list-style-type: none"> • Self-reported levels of poor functioning • Number of hours where participants are able to work per week
Decreased adverse outcomes	<ul style="list-style-type: none"> • Number of alcohol related incidents • Number of domestic violence incidents • Number of suicide related behaviors (gestures, attempts, completed suicides)
Satisfaction	<ul style="list-style-type: none"> • Patient surveys • Focus groups • Number of complaints

One of the simplest and most common ways of assessing outcomes in program evaluation is through the use of self-report measures of symptoms, and there is a plethora of standardized PH and TBI self-report scales. There are clinician-administered rating measures available as well, but these tend to require more time and training, and are therefore used less in general practice. **Figure 23. PH and TBI Instruments** below, provides a listing of a few useful instruments with descriptions of the tools and some parameters for their use. Additional information on other measures can be found in **Appendix VI: Information on Recommended Measures for PH and TBI**, which lists information regarding several measures of PH and TBI that have been recommended for use in military treatment facilities.

Figure 23. PH and TBI Instruments

Tool	Description	Problem Area	Administration Method	Administration Time	Continuum
Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT)	The test battery measures multiple aspects of cognitive functioning following a concussive event. It evaluates attention span, working memory, sustained and selective attention time, response variability, non-verbal problem solving and reaction time.	TBI	Clinician Administered	20-25 minutes	Screen, Assess and Diagnose

Tool	Description	Problem Area	Administration Method	Administration Time	Continuum
Clinician-Administered PTSD Scale (CAPS)	The CAPS is the gold standard in PTSD assessment. The CAPS is a 30-item structured interview that corresponds to the DSM-IV criteria for PTSD. The CAPS can be used to make a current (past month) or lifetime diagnosis of PTSD or to assess symptoms over the past week.	PTSD	Clinician Administered	45-60 minutes	Screen, Assess and Diagnose
Dizziness Handicap Inventory(DHI)	The DHI tool assesses for dizziness, a symptom reported in approximately 80% of TBI cases. The test can be broken down into three parts (functional, emotional, physical) and can be scored.	TBI	Self-Report	5-10 minutes	Screen, Assess and Diagnose
PTSD Checklist – Military Version(PCL-M)	The PCL-M is a 17-item self-report measure of the 17 DSM-IV symptoms of PTSD. The PCL has a variety of purposes, including: screening individuals for PTSD, diagnosing PTSD, monitoring symptom change during and after treatment.	PTSD	Self-Report	5-10 minutes	Screen, Assess and Diagnose

Data Collection Measures- Considerations for Selecting Self-report Measures

Whenever possible, standardized instruments that are considered valid and reliable should be used in lieu of scales created de novo. Using scales that are already validated has several advantages, most notably, that the results can be compared to the outcomes reported from other programs. When programs use non-standardized scales or questionnaires, generalizability of findings is limited.

Self-report tools can have varying degrees of reliability and validity. *Reliability* refers to the ability of an instrument to yield consistent results. There are several subtypes of reliability, including: internal consistency, test-retest, and parallel forms. *Validity* refers to the degree that the instrument is actually measuring what it intends to measure. There are several types of validity, including: content, criterion, and construct validity. **Figure 24. Types of Reliability and Validity** lists what question each of these types of reliability or validity aim to address.

Figure 24. Types of Reliability and Validity

Research Principle	Subtypes	Question
Reliability	<ul style="list-style-type: none"> Internal consistency Test-retest Parallel Forms 	<ul style="list-style-type: none"> How similar are the items within the scale to one another-are they measuring the same underlying construct? Will repeating the measurement yield the same reading? Do two different versions of the scale correlate highly?
Validity	<ul style="list-style-type: none"> Content Criterion Construct 	<ul style="list-style-type: none"> Does the test accurately capture the domain that it is intended to? Does the measure correlate with other valid measures? Does the test result correlate with some external measure of the construct being measured?

Besides the reliability and validity of a measure, one must also consider its appropriateness for the population and planned use. Many measures are tailored to a specific population, and would not be appropriate for use in other settings. Not all measures, even if considered reliable and valid, can serve as outcome measures. For example, the PTSD Checklist (PCL) can be used as a screening/diagnostic tool and as a measure of treatment outcome. The Alcohol Use Disorders Identification Test (AUDIT) is a widely used screening instrument that can assist with diagnostic assessments. The AUDIT is unsuitable for use as a measure of outcomes in substance abuse programs, as the measure has several questions that are historical “Have you or someone else been ever been injured as a result of your drinking?” The answers to such questions will not change due to a 90 day outpatient treatment program for substance abuse, and therefore using this measure would be inappropriate to capture outcomes. The AUDIT-C, on the other hand, is a separate measure consisting of a subset of the AUDIT questions that ask about recent consumption of alcohol. The AUDIT-C would be a much better measure of outcome for an alcohol treatment program compared to the full AUDIT.

Conclusion

A critical component of the program evaluation process is to ensure evaluation questions reflect needs, priorities and relevant program areas. With the right questions, the right information will be collected without the loss of time and money and will provide the opportunity to understand, improve and strengthen the program.

Key Takeaways

- Evaluation questions are rooted in the program’s goals and objectives, its program development phase and the specific purpose of the evaluation
- Developing appropriate evaluation questions is not a single step process and involves program review, developing an evaluation logic model, and determining the most appropriate evaluation type and focus
- Evaluation questions can focus on processes or outcomes. Once developed, evaluation questions must satisfy the SMART criteria
- Collection and measurement tools chosen must be carefully weighed in terms of cost, ease of use, and availability

Sources and suggested readings

- Aubel, J. (1999). *Participatory program evaluation manual*. Calverton: Child Survival Technical Support Project and Catholic Relief Services.
- Chinman, M., Imm, P., & Wandersman, A. (2004). *Getting to outcomes: promoting accountability through methods and tools for planning, implementation, and evaluation*. Arlington: RAND Corporation.
- Fowler, F.J. (2009). *Survey research methods (4th ed.)*. Thousand Oaks, CA: SAGE Publications, Inc.
- Gelmon, S. B., Foucek, A., & Waterbury, A. (2005). *Program evaluation: Principles and practices (2nd ed.)*. Portland: Northwest Health Foundation.
- National Institute on Drug Abuse. (1993). *How good is your drug abuse treatment program?: Overview and case study*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy, Education, and Legislation, Community and Professional Education Branch.
- W.K. Kellogg Foundation. (1998). *Evaluation Handbook [Item #1203]*. Battle Creek, MI: Collateral Management Company.
- Wolfe, B.L., & Miller, W. R. (n.d.). *Program evaluation: A do-it-yourself manual for substance abuse programs*. Albuquerque, NM: University of New Mexico, Department of Psychology.

Step 3: Develop Evaluation Design



After the team has reviewed the program background and has developed and prioritized the evaluation questions, it is time to **Develop the Evaluation Design**, the third step in the program evaluation process.

Developing the evaluation design requires the evaluation team to undergo critical inquiry about what aspects of the program will be assessed, when and from whom data will be collected, and how program performance will be measured. The outcome of this process is the identification of the most appropriate and sound evaluation method(s) for the program of interest.

Evaluation Design

- Develop the Evaluation Design
- Understand Threats to Validity

Evaluation designs are selected in order to construct a sound methodology for data collection that will reduce extraneous variables from impacting the evaluation. By correcting for such variables, evaluation results can be attributed to the program itself, thus reducing uncertainty regarding whether other variables may have caused the observed results. A strong evaluation design reduces the impact that external factors might have had on the outcomes or findings of the evaluation. If the evaluation design is weak and does not mitigate the influence of these external factors, then the evaluation findings will not be viewed as valid.

Developing an evaluation design takes time, demands particular attention to detail, and necessitates informed decision-making. The choices made in this step will shape the framework used to carry out the evaluation, thus greatly affecting the legitimacy of evaluation results.

Develop the Evaluation Design

In Step 2, the evaluation questions were selected and finalized, data was identified for collection, and tools were selected to capture the data. There are three main categories of evaluation designs: descriptive, experimental, and quasi-experimental. Each type has its own strengths and limitations, and selection depends on the nature of the program and factors such as cost, level of expertise needed, and whether the evaluation will be classified as research and require Institutional Review Board review and approvals. The following paragraphs provide an overview of the three main categories, including examples of each.

1) Descriptive Designs

Descriptive evaluation designs are utilized to provide a picture of a program or some aspect of a program. This design is best used to analyze program performance by collecting feedback on services and outputs, and is therefore more likely to be used if conducting a process evaluation. Compared to the experimental designs described next, descriptive designs are lower in cost and easier to implement, and can produce results in shorter periods of time.

Descriptive research designs do not utilize random assignment of participants into groups, or have separate intervention and control groups. This type of study design is therefore unable to answer questions of cause and effect. Results from descriptive evaluations can be used to refine program processes, goals and objectives, as it can provide information describing the manner in which the program operates (e.g., is the program serving the intended population), and can render feedback about the services being provided (eg., satisfaction).

Examples of descriptive designs variables:

- Patient satisfaction survey that seeks to determine how happy patients are with a particular aspect of the program, and their willingness to participate again
- Examination of clinic appointment data to understand if newly hired providers are being fully utilized
- Examination of patient demographics to determine if the attendees in a program intervention align to the expected population

Experimental Study Designs are used to control for extraneous variables that may threaten the validity of the evaluation results, and can be classified into experimental or quasi-experimental designs. These designs will be represented in the descriptions below in diagrams using the following notation:

R = Random assignment into a group
 O = Observation of the program result
 X = Intervention

2) *Experimental Design*

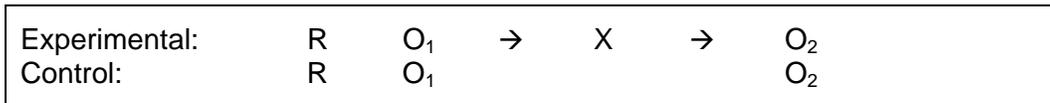
Experimental studies seek to test the relationship between two variables, specifically whether an independent variable which is manipulated has measurable effect on a dependent variable. This type of study aims to determine cause and effect relationships. In the case of program evaluations, the interventions or activities a program produces (e.g., therapy sessions, medications, classes) are the independent variables and the dependent variables are the expected outcomes (e.g., reduction in symptoms, return to duty) of the program. For an evaluation design to be *experimental*, it has to satisfy two criteria. 1) Program participants are randomly assigned to either an intervention group (a group that receives a treatment, intervention or program service) or to a control group (a group that does not receive any treatment or service). The control group should be equivalent to the intervention group in all relevant demographic respects; such as age, gender, economic status. A control group is established in order to provide a basis for sound comparison, and determine if the intervention provided was the actual cause of the outcomes observed. 2) The program must render a treatment, modality, service, workshop or other intervention, within a defined timeframe.

Random assignment of participants into an intervention or a control group is the hallmark of experimental designs. Randomization seeks to assure that each participant has an equal chance of being assigned to either group, thus minimizing bias. With randomization, evaluation findings that show a change in the intervention group can be attributed to the program.

Experimental evaluations may be costly, time consuming, and challenging to administer. The majority of program evaluations will not utilize a true experimental design, due to costs, time and difficulties in establishing control groups or randomizing subjects. Another significant impediment to using experimental designs lies in the fact that these designs clearly fall under the category of research and thus require very stringent approvals and oversight from an Institutional Review Board. It is for these reasons that the majority of program evaluations will use one of the quasi-experimental designs described below.

Experimental Design Example: One hundred patients with anxiety disorder who currently attend an outpatient clinic are randomly assigned into an intervention and a control group. Those in the intervention group attend one counseling session per week for six months, for a total of 24

counseling sessions. The control group is not provided with counseling sessions. Both groups have similar characteristics. Prior to the start of the counseling sessions, both groups undergo a standardized psychological assessment instrument (O_1). The intervention group completes all sessions (X), while the control group does not undergo this counseling. Once the intervention period is complete, both groups are given the same psychological assessment instrument (O_2) to assess whether there is a significant reduction in anxiety levels between the intervention and the control group.



3) Quasi-Experimental Design

A quasi-experimental design is very similar to an experimental study, in that they both seek to establish the existence of a cause and effect relationship between an intervention and a particular outcome. The primary difference between a true experimental design and quasi-experimental design is randomization of group assignment. Quasi-experimental designs are used when randomization of subjects is not possible. In certain instances, withholding program services may not be ethical, or is impractical based on the structure and functioning of a program. In these instances, control groups cannot be created. It is sometimes possible to create comparison groups, which are often called non-equivalent comparison groups, to indicate they were created via randomization. The following section presents several quasi-experimental study designs that are often used in program evaluation.

A *One Group, Post-Test Only Design*, also known as the one-shot design, is a type of quasi-experimental evaluation whereby participants are all placed into one intervention group. There is no randomized assignment of participants into groups, and no control group, thus no meaningful comparisons can be made. The intervention group is given an intervention (X), and in some future time, data about the intervention effects are collected using a post-test (O). This type of design is the weakest of all, since the data obtained from a post-test does not provide sufficient information about patient performance, patient improvement, nor can it establish that the program was the cause of any noted changes to patients. This type of design can be valuable in providing pilot data applicable for future evaluations.

One Group, Post-Test Only Design Example: All PTSD patients who currently attend an outpatient clinic attend a one-day long course to learn several relaxations techniques to reduce stress (X). They are taught to use five different evidence-based techniques during this intervention. A month later, they are provided with a structured post-test questionnaire (O) to determine their levels of anxiety.

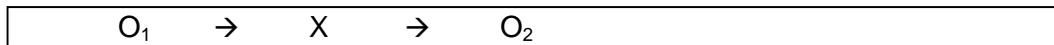


The *One Group, Pre-Test/Post-Test Design* is very similar to the above design, as it is a quasi-experimental evaluation where all participants are non-randomly assigned into one intervention group, and no control group exists. The major difference is that a pre-test (O_1) is given before the program services begin to establish a baseline before the intervention is provided (X). A post-test (O_2) is administered to gather data about the results of the program after the intervention period. This design provides information about the changes that patients have experienced between the time of the pre-test and post-test.

While this design is stronger than the One-Group, Post-test only design since there is no comparison group, it is not possible to conclude whether any changes noted between the time

of the pre-test to the post-test are specifically due to the program. Perhaps other outside variables that were unaccounted for led to the observed change, such as environmental events. There is no certain way to know if other factors are a cause for the results observed.

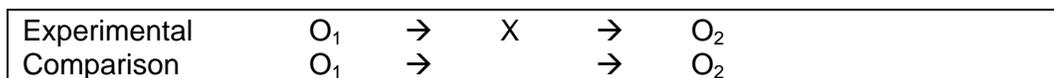
One Group Pre-Test/Post-Test Example: All PTSD patients who currently attend an outpatient clinic take part in a one-day long course to learn stress-reduction techniques. Before starting the course, they are given a short, pre-test measure of anxiety. A month later, they are provided with a structured post-test questionnaire, asking questions similar to the pre-test, to determine if they experienced a reduction in anxiety levels.



The *Non-Equivalent Comparison Group Design* is also a type of quasi-experimental design whereby program participants are non-randomly divided into an intervention and a control group. Prior to providing an intervention, both groups are given a pre-test (O_1). In some future time period, the intervention group is given an intervention (X), while the control group does not receive any intervention. Once the program intervention is complete, both groups complete a post-test (O_2).

Since there is no randomization of group assignment, the two groups may not be similar in terms of important characteristics (such as familial psychiatric history, or a disproportionate number of individuals within a specific socioeconomic category), thus the name non-equivalent comparison group. The feature of non-equivalence introduces extraneous variables that can interact with program services. The results obtained from the program evaluation may not be credited to the services or intervention provided by the program, which makes it difficult to conclude with certainty that the outcomes measured were not caused by other factors.

Non-Equivalent Comparison Group Design Example: All patients with PTSD who currently attend an outpatient clinic are offered a one-day long course to learn stress-reduction techniques. Those who do not want to attend the course form the non-equivalent comparison group, while those who do attend will be the intervention group. Prior to the course, both groups answer a pre-test questionnaire measuring their levels of anxiety. A month later, both groups are asked to complete the same measure of anxiety as a post-test.



Control and Comparison Groups

There are two general types of groups, used to help evaluators isolate treatment effects-- control and comparison groups. Control and comparison groups are similar in that participants within these groups complete the same measures as the group receiving the interventions, however they do not actually participate in the interventions. They should also be similar to the group receiving the interventions on key variables. The key difference between a control and a comparison group lies in the fact that a control group is formed by randomly assigning participants, while a comparison group is formed in a non-randomized manner.

Because random assignment is not always possible, utilize control groups is oftentimes not possible; however, there may be a suitable comparison group available. Some common sources for comparison groups include:

- Participants who are on a wait list to be enrolled in the program

- Participants who were unable to attend some key part of a program (e.g., being seen for medication management, but unable to attend counseling)
- Participants who are being seen at a similar type of program

Designs that add a control or comparison group are much stronger than designs without them. Control groups allow for isolation of the treatment effects, which allows for confidence that positive effects from the program are in fact due to the programs activities / interventions. Designs without a control or comparison group cannot rule out that the effects seen are due to some other factor. Such designs are still valuable, in that they can show that participants improved on measures during the timeframe of the evaluation, and lay groundwork for future evaluations.

Whether or not the evaluation will be considered research is important to decide before selecting the design for the program evaluation. Normally simple program evaluation is not considered research; however, there are several things that will automatically move an evaluation into the category of research. The use of randomized groups is normally considered a key facet of research, and if the design randomizes participants into groups, it will likely be classified as research. Forming a comparison group for the purpose of the evaluation (starting a wait list as opposed to using an existing wait list) would also be considered a research activity.

The evaluation plan may need to be submitted to a local institutional review board (IRB). IRBs are established to ensure that research involving human subjects ensures minimal risk, which is defined by Title VII of the Code of Federal Regulations as “the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.” An IRB may review the evaluation design and determine that it does not qualify as research, in which case the IRB will grant an exemption to formal IRB monitoring. When there is question about whether a program evaluation needs to be submitted for approval or exemption, it is best to seek the IRB point of contact to obtain guidance.

Understanding Internal Validity

It is important to understand how internal validity is related to the study design. *Internal validity* addresses whether there is a causal relationship between a given intervention and some measured outcome. Demonstrating changes in outcome variables is important, but being able to identify specific interventions or program initiatives as the cause of these changes is even more important. In order to establish and maintain good internal validity, program evaluators should seek to minimize the impact of recognized threats to internal validity. Flaws in the design of the evaluation can compromise an evaluator’s ability to determine the specific causes of any changes measured in outcome variables.

There are four general ways to minimize threats are described below:

1. Standardize the conditions under which the evaluation study is conducted.
2. Gather detailed information about evaluation participants.
3. Collect detailed information about the procedural details of the research study, for example: where and when the study occurs.
4. Utilize an appropriate research design.

Figure 25. Threats to Internal Validity provides common threats to internal validity, their impact on findings, and methods commonly used to control for each of these threats.

Figure 25. Threats to Internal Validity

	Threat Type	Description	Affect on Validity	How to Minimize
Threats Related to Equivalence of Groups	Statistical Regression	Patient subgroups with extreme scores may show movement towards the population mean, since their scores were extreme to begin with.	Scores become less extreme due to normalization, and not because of the intervention.	Avoid using extreme scores.
	Attrition	Uneven drop-outs in different subgroups within the study, leading to non-random distribution of participants between groups.	Leads to a biased post-test score.	Use large groups sizes when possible.
	Selection Bias	Bias introduced into the evaluation during the group assignment process.	Evaluation findings are incorrectly attributed to the program because the intervention group showed a change.	Use random selection if possible; Use statistical control procedures if groups are not randomized.
Threats Related to Controlling of Extraneous Factors	Maturation	This occurs when patients change across time, sometimes due to physical development, or improvement over time without any intervention.	Change is attributed to time, and not to the intervention.	Use a control group if possible; Minimize length of study if possible.
	History	Unforeseen environmental factors that occur between the pre- and post-test, especially if this occurs to only one group, which influence participants' outcome measures.	Unable to conclude that only the program activities impacted the evaluation findings.	Use a control group if possible; Minimize length of study if possible.
	Testing	Repeated testing using the same types of tools or questions can affect the results of the measure.	Difficult to determine if patients' answers were influenced by their familiarity with the instrument.	Use parallel forms for measures if possible.
	Instrumentation	The data collection tool is not employed consistently during repeated measures, such as when data collection personnel change their rating criteria.	Outcomes measures are gauged inconsistently.	Standardize administration of measures.
	Hawthorne Effect	Effect of knowing that one is being observed changes behaviors	Participants change their behaviors simply due to being observed, rather than a real effect of the intervention	Minimize the degree of overt attention by using unobtrusive measures when possible.

Conclusion

Development of the evaluation design is a critical and time-consuming step in the program evaluation process. It requires evaluators to make decisions regarding the specific aspects of the program to be assessed, when and from whom data will be collected, and how program performance will be measured. Making these decisions early in the planning process will provide the basis for a strong data gathering and analysis plan and will reduce the risk of collecting irrelevant information.

Key Takeaways

- The design of the evaluation depends on the questions to be answered, the population to be studied, and the data available for collection
- Evaluations can be strengthened by understanding the potential threats to internal validity, and accounting for them whenever possible

Sources and suggested readings

Chinman, M., Imm, P., & Wandersman, A. (2004). *Getting to outcomes: Promoting accountability through methods and tools for planning, implementation, and evaluation*. Arlington: RAND Corporation.

Cook, T. D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.

Cornell University Law School. (n.d.). *7 CFR 1c - Protection OF Human Subjects*. Retrieved May 11, 2012, from Legal Information Institute: <http://www.law.cornell.edu/cfr/text/7/1c>

National Institute on Drug Abuse. (1993). *How good is your drug abuse treatment program?: Overview and case study*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy, Education, and Legislation, Community and Professional Education Branch.

U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). (2001). *Guidance for industry E 10 choice of control group and related issues in clinical trials*. Rockville, MD: Author.

W.K. Kellogg Foundation. (1998). *Evaluation Handbook [Item #1203]*. Battle Creek, MI: Collateral Management Company.

Step 4: Develop a Data Plan



Once the evaluation team has decided on an appropriate evaluation design, the final step in the Preparation Phase is to **Develop the Data Plan**. The purpose of this step is to codify the procedures that will be used to gather, store and analyze the data collected for the program evaluation. It is important to explicitly outline these steps prior to the actual data collection, as a good data plan can prevent redundancy of efforts, confidentiality or legal breaches, and can generally help to minimize the costs and efforts associated with program evaluation.

The sections that follow will walk evaluators through a step-by-step process to ensure their data are collected, secured, and analyzed correctly. It is recommended that the steps be followed in the order presented to ensure that required protocols and procedures are in place before data collection begins.

Data Plan

- Develop Data Sampling Plan
- Develop Protocols for Securing Data
- Determine Data Analysis Plan
- Develop Standard Operating Procedures
- Obtain Appropriate Approvals
- Develop Database
- Conduct Staff Training and Pilot

Developing the Data Sampling Plan

The data sampling plan provides a detailed protocol for all aspects of data collection. Steps to consider when developing this plan include:

Identify the sample to be studied. Who will serve as the evaluation study sample? Again, depending on the purpose of the evaluation and the evaluation questions to be addressed, samples may differ widely. Program staff, clients, patients, family members, etc. may all be feasible study samples depending on the evaluation questions to be studied. It is important to identify the sample early in the data sampling plan to inform the next several steps.

Identify data already being gathered. Is there any data that is already being collected? If so, is this data useful? If there is already information that has been collected, it will prevent unnecessary work and time collecting data. Even if a program is collecting some data during the intake process, one should remember that not all measures are able to serve as outcome measures (e.g., AUDIT vs. AUDIT-C).

Identify the type of data to be collected. Most likely, evaluators have already started to consider the type of data they will collect during **Step 2: Develop Evaluation Questions and Step 3: Develop Evaluation Strategy** when developing evaluation questions and study design. Decisions regarding the type of data to collect (e.g., qualitative and/or quantitative data, self-report data or observational data, etc.) will also help to inform the next step of the data sampling plan.

Identify the procedures and instruments to be used during data collection. Once the sample and type(s) of data to be collected have been identified, evaluators should be able to determine the procedures and instruments required to complete data collection. Initial decisions regarding this step also will likely have been discussed during **Step 3: Develop Evaluation Strategy** as the study design may dictate specific data collection sources or

procedures. The process of identification of procedures and instruments is likely to be affected by budgetary concerns as well. Certain measures may be protected and require licenses or fees to procure. Special training may be required to administer or score some instruments, which leads into the next step of the data sampling plan.

Identify who will collect, score and enter the data. The procedures and instruments to be used during data collection will dictate, in part, the skills and proficiencies required of the program evaluation study staff. Do staff members need to be trained in specific procedures related to data collection, including how to conduct certain data collection procedures, administer data collection measures, score instruments or output from procedures, enter the data into a database, etc.? These questions should be considered well in advance of data collection to prevent strains to the budget and timeline of an evaluation.

Template L: Data Sampling Plan
(double click on image to open)

Identify the timeline for data collection.

Once the pieces described above are in place, evaluators should have the information necessary to construct a timeline for the data collection process. When projecting a timeline, remember that once data have been collected from the sample, steps such as scoring and entering of the data may add significantly to the total amount of time required for this step.

If the evaluation will include any follow-up with participants who are no longer in the program, there are additional logistical requirements. Specific information about collecting follow-up information is discussed in **Appendix VII: Conducting a Follow-Up Evaluation.**

Evaluation Question	What Data Will Be Collected?	How Will Data Be Collected?	From Where Will Data Be Collected?	When Will Data Be Collected?	Who Scores Measure? (If Applicable)
EXAMPLE: Are our patients with an Axis I diagnosis of depression (MDD or Dysthymia) showing clinical improvement?	Beck Depression Inventory-II (BDI-II)	Paper and Pencil Self-Report	<ul style="list-style-type: none"> All patients in program Any patient with depressive disorder 	<ul style="list-style-type: none"> Upon entry to program as part of intake packet At the end of treatment or termination session 	<ul style="list-style-type: none"> Technicians trained in scoring BDI-II must be reviewed by provider (due to suicide question)
1.					
2.					
3.					

A data sampling plan worksheet can be found in **Template L: Data Sampling Plan.**

Develop Protocols for Securing Data

During the program evaluation, it is likely that sensitive data will be involved. As with all other types of research and data collection, ensuring the security of the data is crucial to maintaining the privacy and confidentiality of the study sample and should be a top priority in any data plan. For programs using identifiable health information or patient data, procedures to assure compliance with the Health Insurance Portability and Accountability Act (HIPAA) are critical as well.

The term *personal identifiable information (PII)* is commonly used to refer to information about an individual that is maintained by an agency and may include their name, address, photograph, and/or information about personal characteristics. One method for ensuring patient or subject confidentiality is to de-identify such sensitive information. *De-identified information* requires that information about an individual has been removed so that it cannot be connected back to the particular person or source from which it came. This is a common security technique used in research to protect the confidentiality of subjects and patients. The level of de-identification possible for a study will depend on the type of data necessary to answer the evaluation questions. Demographic information such as age, sex, rank, etc. may be very informative to the study but its inclusion should be considered carefully if there is a chance that those data could be traced back to specific individuals.

Data protocols and standard operation procedures (SOPs) serve as a framework to guide data management staff through required processes, including security measures. These approaches and guidelines enable evaluators to maintain secure data management systems at all times. Evaluators must establish an appropriate infrastructure for data storage and security measures, as well as explicit guidelines regarding who can access the data, well in advance of data collection.

Paper or hard copies of data, to include surveys, patient charts, etc., should always be kept in secure filing cabinets and both the cabinet and the room in which it is housed should be secure - a process known as “double-locking”. It is also important to determine the required duration of data housing. Many institutions or organizations have explicit instructions regarding how long hard copies of data must be retained as well as specific instructions for destruction of these

documents once the specified time has passed. Evaluators need to be familiar with these procedures before data collection begins.

Electronic data sources, including electronic health records (EHR), databases containing study data, etc., must also be secured using specific measures. Electronic resources should exist in encrypted or password-protected databases and should never be transmitted electronically without proper encryption techniques in place. Many institutions or organizations have specific training for individuals who will have access to this type of data. Evaluators must ensure that access to electronic databases is limited to those who have undergone the proper screening and training, and who have a valid reason for such access.

DATA SECURITY CHECKLIST

- Review data security requirements of the sponsoring institution
- Review HIPPA requirements if using identifiable patient or health data
- Establish procedures for de-identifying PII
- Develop SOPs describing data security measures, including:
 - Personnel qualifications and restrictions for handling data
 - Regulations on copying data
 - Procedures for transporting or transmitting data
 - Procedures for storing data, during and after the evaluation
 - Procedures for destruction of data (and copies)

It is advisable that all evaluators familiarize themselves with the specific security requirements of their individual organization as these procedures can vary. The program evaluation team should collaborate with data management staff to develop and customize a data security checklist.

Determine Data Analysis Plan

Although the analysis of evaluation data falls under **Step 6: Analyze Data** in the program evaluation process, it is advisable to have a data analysis plan in place *before* any data collection takes place. This will allow evaluators to plan appropriately and address issues such as sample size, requirements for statistical software and data format, whether the services of a statistician or analyst will be required, etc.

For specific information regarding formulating a Data Analysis Plan, it may be helpful to read ahead to the section describing **Step 6: Analyze Data** where information on choosing an appropriate statistical test and software package is presented.

Develop Standard Operating Procedures

The development of SOPs is strategically located at this point in **Step 4: Develop Data Plan** to allow for a well-informed plan as well as a plan that can provide guidance for the next

several steps of the program evaluation process. Now that important information such as the process for data collection, storage, security and analysis has been identified, these steps can be formalized in a document that will serve as the framework for the data plan moving forward. Protocols provide staff with a written guide to follow and consult when questions arise. Well-defined protocols and SOPs provide consistent direction, reduce training time, and improve work consistency throughout the program evaluation.

SOPs are particularly relevant for program evaluations that will employ a large staff or will require the efforts of multiple individuals doing the same or similar tasks. Consistency in procedures is absolutely critical in ensuring a valid and reliable evaluation. The SOP will also help to inform staff training by identifying the processes and skills necessary to complete the data collection, storage, security and analysis process.

Using information from the previous steps, identify the key areas of concern where standardized instructions would be most useful. This will likely include procedures for collecting data, including how to administer certain instruments or scripts for interacting with subjects or participants, procedures for the secure storage and handling of collected data and the basic process for data analysis. In addition, SOPs should include information on staff requirements, such as procedures for obtaining clearances or training required for data collection or use. The more detailed and explicit the SOPs, the less likely evaluation staff will be to make mistakes or have to address uncertainties in procedures during the actual evaluation process. Although SOPs are not completely static documents and may change slightly based on the next several steps in the Data Plan, efforts should be made to have complete and detailed SOPs in place before data collection begins.

Obtaining Appropriate Approvals

The specific requirements for approvals will likely differ based on the institution in charge of the program evaluation or the actual location of the evaluation. Evaluators must familiarize themselves with the guidelines, rules, and laws (if applicable) governing program evaluation within their organization. Information about required approvals and policies will often be found in an organization's Research or Performance Improvement offices. Staff in these departments will be familiar with requirements for obtaining various authorizations before beginning data collection. For programs within the Defense Department, evaluators should also obtain approvals from the Chain of Command.

Most program evaluation protocols qualify as IRB exempt; committees that were established by federal regulations to review and approve any research involving human subjects. However, each IRB has specific requirements and instructions for obtaining exemption status. Any organizations that support human research have an IRB that should be consulted. Although a review may not be required, evaluators may still need to submit paperwork establishing exemption status. It is very important to verify IRB requirements and procedures before proceeding with the evaluation process.

Developing a Database

The development of a database to house evaluation data will be affected by many of the previous steps in the Data Plan. When choosing or developing a database, evaluators should consider:

- The type and amount of data to be stored in the database
- How data will be entered or imported into the database
- Security options for the database

- Compatibility of the database with any statistical procedures to be completed or statistical software to be used

There are many options to consider when choosing or designing a database. One of the largest distinctions between types of databases are flat-file versus relational. A *flat-file database* uses a single table as its basis. In flat-file databases, evaluators can establish categories as well as individual subject entries but these data cannot be shared or merged across other tables or databases. One commonly used example of a flat-file database is *Microsoft Excel*. On the other hand, *relational databases* incorporate multiple data tables and have established procedures for the tables to merge or interact. These may allow for the sharing of data across networks, via the internet, etc. A commonly available example of a relational database is *Microsoft Access*. Each type of database has associated benefits and challenges, so it is important to consider database options carefully.

In choosing between a flat-file database and a relational database, foresight must be applied into what will be done with the data. Flat-files do not allow for automation, which means that if there are two or more flat-files containing patient information and a patient's allergies change, the allergies would need to be manually modified in each file. However, flat-files are capable of storing information, and allow for printing, manipulating fields, and formatting information. A flat-file database is a simpler way to store data as long as the data will be single file in the form of rows and columns without relationships or links between records. A relational database incorporates multiple tables allowing them to work together. This supports comparison of relationships and merging and displaying of all information across tables. If multiple tables will need to be compared, this may be a safer, more judicious choice that will not cause reentering of data or conversion to a relational database later when such comparisons or merging of tables is needed. Relational databases are more robust in terms of capabilities, but this comes at a cost of being more complex to create and maintain.

Both flat-files and relational databases offer data-validation options that should be utilized in data entry. Microsoft Access has an "input mask" feature that will require data entry into database fields to conform to some pre-determined rule. For example, forcing rank to be entered as a letter then number (i.e., E4, E7, O3 and O6) will increase consistency in how data are entered, saving time and effort. Microsoft Excel has the ability to create data validation rules.

Once a preferred database type has been identified, the database can be prepared for data import or input prior to data collection. Information on the amount and type of data should be available given earlier planning steps so the database can be designed and labeled in preparation for data collection. This will facilitate the rapid availability of data for analysis once collection has been completed.

Conduct Staff Training and Piloting

Some elements of staff training may have already been addressed in the previous step, *Obtaining Appropriate Approvals*. Certain approvals may require documentation of staff training for issues such as data collection, patient privacy, and information assurance. If evaluation staff have not yet completed these institutional level trainings, it is important to complete this step before commencing data collection. Ensure that all staff members are compliant with all applicable institutional, federal, and DoD laws mandating data collection. Depending on the anticipated timeline for the evaluation, it may be necessary to renew training certifications so staff and evaluators should be aware of all such deadlines and requirements.

As with SOPs, consistent training procedures will ensure that all staff conducts data collection processes in a similar manner, thus reducing variability and error in the program evaluation procedure. If any of the procedures of instruments to be used for data collection require specific training, as identified when *Developing the Data Sampling Plan*, this is the point in the process to provide such training for staff. Evaluators may want to consider administering proficiency checks before the data collection begins in order to verify consistency.

Evaluators may also want to pilot certain procedures or instruments at this point in the development of the Data Plan. This is particularly true if the instruments or measures to be used are new or require specific or precise skills for administration. Piloting allows for the early identification of potential problems or issues so that they can be addressed before the actual program evaluation commences. In addition to piloting certain procedures or instruments, evaluators in charge of conducting very substantial or large-scale evaluations may want to do a small-scale pilot of the entire study as a way to verify feasibility before working with a much larger sample.

Conclusion

Developing a data plan is the final step of the Preparation Phase of the program evaluation and becomes the blueprint for the Execution Phase. It is important to include complete and detailed information on the data to be collected, the methods of collecting and safeguarding data, and confirmation that the appropriate approvals have been obtained. Significant time and effort should be devoted to this step as it provides the majority of the framework and quality checks for data collection and analysis.

Key Takeaways

- Advanced planning is the key to a well-executed program evaluation
- Formalizing procedures and plans for data collection, storage, security and analysis in SOPs before data collection commences is imperative to the evaluation success
- Evaluators and staff should be well-versed on the appropriate approvals and trainings required to perform data collection
- Training and piloting for procedures and instruments will facilitate the process of data collection

Sources and suggested readings

Gelmon, S. B., Foucek, A., & Waterbury, A. (2005). *Program evaluation: Principles and practices* (2nd ed.). Portland: Northwest Health Foundation.

National Institute on Drug Abuse. (1993). *How good is your drug abuse treatment program?: Overview and case study*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy, Education, and Legislation, Community and Professional Education Branch.

Olney, P. C., & Barnes, M. S. (2008). *Collecting and analyzing evaluation data: Planning and evaluating health information outreach projects booklet 3*. Bethesda: National Library of Medicine, Outreach Evaluation Resource Center.

W.K. Kellogg Foundation. (1998). *Evaluation Handbook [Item #1203]*. Battle Creek, MI: Collateral Management Company.

Phase 2: Execution



The Execution Phase represents the second phase of the program evaluation process. The preparation steps completed during Phase I were utilized to develop a Data Collection Plan. This plan will provide a framework to guide evaluators through the subsequent implementation of that plan in the next steps of the program evaluation process. The Execution Phase consists of two steps: **Step 5: Gather Data**; and **Step 6: Analyze Data**.



Step 5: Gather Data



Implementing the Data Plan

Once the evaluators have refined the questions, determined the appropriate evaluation methods, and developed the data plan, the data gathering process can begin. This is the point in the process where the evaluation team will implement the data collection plan. Selecting quality, reliable data is essential to program evaluation, and ensuring quality data gathering is essential to minimizing threats to validity, and producing reliable, and informative program data.

Gather Data

- Implement the data plan
- Develop and execute a data monitoring plan
- Check data for validity and reliability
- Secure the data according to the data plan

An effective data plan ensures a viable collection process. This plan, written during **Step 4, Develop Data Plan**, summarizes the various types of data and describes the methods for designated staff to use to obtain data. Refer back to the information covered under Step 4 of this guide for additional information. Tasks associated with the data gathering process can be performed by various qualified team members. Lead evaluators should assign roles based on each team member’s experience, qualifications, and professional training and work load.

Develop a Monitoring Plan

A data monitoring plan further identifies the process by which data is collected, entered, validated and stored by key personnel. Throughout the data gathering process it is important for the evaluation team to frequently review the plan to ensure that accuracy is continuously maintained, data quality is maximized and to ensure secure handling and storage of collected data. A sample template for this plan is provided in **Template M: Data Monitoring Plan**.

The Data Monitoring Plan should address procedures for collecting, reviewing and safeguarding all data collected step. All staff assigned to handle data need to be trained prior to undertaking this step. Training should include the instructions given to patients or clients from whom data are collected, the method of collection, procedures for recording and handling the data, and how to safeguard data collected. An individual from the evaluation team should also be assigned to verify compliance with the data plan and the data monitoring plan throughout the collection process.

Collecting Data. Specific staff should be designated to collect and handle the data. This is important in order to protect the integrity of the data and to ensure consistency throughout the collection process.

Data collectors are also in the best position to monitor how the process is working and how it is received by the people from whom the data is received. They can provide feedback to the evaluation team if changes

Template M: Data Monitoring Plan
(double click on image to open)

TEMPLATE M. DATA MONITORING PLAN

This document serves as a template to help guide evaluators through the program evaluation process. The data monitoring template is supplemental to the data collection plan and the sampling strategy template. This document is intended to support the establishment of a quality data monitoring plan, and to assist with the collection, documentation, of quality data. This template provides a framework to assist data collection staff and evaluators to ensure the collection and maintenance of quality, valid and reliable data.

DATA MONITORING PLAN
<p>1. Types of Data The review of this data on a consistent basis will allow for the overall understanding of the quality of data monitoring and verification checks.</p>
<p>2. Roles and Responsibilities Evaluators should designate the specific roles and responsibilities of key data collection staff. Consider the following questions:</p> <ul style="list-style-type: none"> • Who will verify data accuracy, by what method and how frequently? • Who will verify compliance with the program plan? • How will compliance be verified and how often? • Will a data monitoring committee be formed? Describe the committee (if applicable, include name, credentials, title, organization and contact information of each member). • What are the mechanisms for maintaining independence of judgment?
<p>3. Monitoring Schedule Develop a schedule that identifies when data will be reviewed, assessed, and verified (i.e. monthly or quarterly).</p>
<p>4. Criteria Identify the criteria that should be examined to assess data accuracy and validity.</p>
<p>5. Action Steps Specify the sequential steps that staff should follow to correct any data inconsistencies.</p>
<p>6. Reporting Develop processes and instructions by which data management staff will perform quality checks in preparation for quality reviews. Detail the process for documenting and reporting and tracking any data inconsistencies.</p> <ul style="list-style-type: none"> • What are the timeframes for reporting? • What mechanism will be used to report (specify forms and procedures)? • Who will prepare and submit the report? • How will outcomes be communicated to stakeholders?

need to be made. The data monitoring plan should include procedures for gathering feedback on the data collection process.

Reviewing Data. Before data is entered in a database, the data must be reviewed for accuracy. The data monitoring plan should specify who is assigned to verify the accuracy, by what method, and how frequently. Ideally this should be done by someone who is not involved with the data collection, particularly for large collection efforts. A data monitoring committee, whose role is to receive regular reports from the collection team or designated data monitors should be formed. This committee can monitor, identify and address data flaws as needed.

Reviewing the data should also include checking the data for validity and reliability. Refer back to **Step 3, Develop Evaluation Design** for information regarding valid and reliable data.

Collected data will need to be entered in a data file. The plan should address where and how it is entered. Methods might include written entries, computer database entries, or a manual filing system. This section of the monitoring plan will also include procedures for making quality checks on procedures the program has in place for data coding, scoring and storage. Scoring and coding also covers procedures to detect coding errors that could undermine the validity of the evaluation results.

Safeguarding Data. Keeping data confidential is vital to any program involving patients. *Personally identifiable information (PII)* must be protected or excluded from the data collected. Such information includes names, addresses, or any other information that might disclose a specific individual's identity.

The monitoring plan should also specify where the data is stored. If data is collected both in paper form and electronically, both must be addressed at this point in the process.

Another important consideration during the data collection process is to ensure that only data which is needed to answer the question is collected. A common pitfall of data collection is collecting information that is superfluous to the study. This can impede and complicate the process of reviewing the data requiring additional time to sort and review unnecessary information. An important point to bring up during training is to "collect only the information you are going to use, and use all the information you collect" (Kellogg, 1998).

Conclusion

The Gather Data step implements the Data Plan developed during Step 4 and requires that evaluation team members be properly trained in the collection and handling of data. This process should be monitored continuously. Executing a detailed Data Monitoring Plan will ensure that everyone involved in data collection performs each step in a specified and consistent manner, the data is reviewed for accuracy and that data is safeguarded throughout the process.

Key-Takeaways

- Develop a Monitoring Plan that will outline specific, step-by-step procedures to gathering data
- Check data for validity and reliability
- Collect only the data that is needed
- Continually monitor collection procedures and safeguard collected data

Avoid These Data Gathering Pitfalls

- Data collectors not trained
- Inconsistent collection techniques
- Unnecessary data collection
- Not safeguarded to protect individual privacy

Sources and suggested readings

Olney, P. C., & Barnes, M. S. (2008). *Collecting and analyzing evaluation data: Planning and evaluating health information outreach projects booklet 3*. Bethesda: National Library of Medicine, Outreach Evaluation Resource Center.

Step 6: Analyze Data



The goal of **Step 6, Analyze Data**, is to examine the data gathered in the previous step using a variety of techniques. Analysis leads to conclusions pertaining to the evaluation questions. The process of data analysis includes the inspection, cleansing and analysis of the data and interpretation of the results. The following section describes the steps in this process while providing examples relevant to psychological health program evaluation.

Analyze Data

- Qualitative and quantitative analysis
- Data preparation using descriptive statistics
- Using inferential statistics to examine relationships
- Interpreting results and drawing conclusions

If the evaluation team does not consist of individuals who are proficient in statistical analysis, it may be advisable to gain permission from program leaders to consult with statistical professionals during this phase of program evaluation. This guide provides only a fundamental overview of data analysis, but evaluation and analysis beyond basic statistics may require the guidance and input of a trained statistician or analyst.

Qualitative and Quantitative Analysis

Program evaluation data will most likely fit into one of two categories: qualitative or quantitative data. Depending on the measures being used, data of both types may be collected from a single program evaluation. It is important to identify the type of data being reviewed, as this will have a major impact on the techniques and tests necessary to analyze the data.

Qualitative data refers to descriptive or characteristic data, most often collected through techniques such as interviews, observations, journaling, etc. It is difficult to quantify or simplify these data to a numerical value because this type of data often addresses abstract concepts such as emotions, themes, or styles. As such, qualitative data may present a challenge to the data analysis process which is often empirical or statistical. Despite the less empirical nature of qualitative data, there are specific techniques for the analysis of this type of data. For example, evaluators can use the technique of categorization to identify recurring or common themes or responses across interviews or open-response surveys. On a more global level, contextualization techniques allow evaluators to examine the relationship between identified categories or responses to prevent over-simplification of these data. **Appendix VIII: Guide to Coding Qualitative Data** provides information on these techniques.

Quantitative data and analyses are those most often associated with statistical analyses because of the numerical or empirical nature of the data. These data quantify *something*: number of appointments, satisfaction ratings on a scale of one to five, total score on a measure of anxiety, etc. The majority of this guide discusses techniques and methods for analyzing quantitative data. The process of analysis may be somewhat intimidating as it often involves statistical tests and software. Consultation of a statistician or analyst may be helpful during this process. However, as this text describes, there are many options that evaluators may use to describe, summarize, and interpret data that are accessible at all levels of statistical and analytical knowledge and experience.

Inspecting and Cleansing Data

Before conducting any statistical analyses, it is critical to inspect the data. This initial inspection allows for the identification of redundant or unnecessary data and cleans and reviews data that may have been entered or processed incorrectly. Errors in data entry can present significant problems during the statistical analysis if they are not identified early in the process. Missing or aberrant data points can significantly skew or change the results of statistical tests, making this step critical to the correct analysis and interpretation of program evaluation data.

Specific study designs may require particular considerations during this phase. For example, longitudinal studies with multiple data points per subject benefit from review of each data set for completeness. Questions to consider might include: Did all subjects or participants provide data for each of the required time points or are there missing and incomplete profiles? Are there particularly discordant data points within the profiles that may indicate an error in data entry? It is also important to remember that in small studies a single outlier or data entry error may result in a significant distortion of the results; hence, extra care should be taken to verify the data have been entered correctly. Evaluators should always be aware of the potential for bias when inspecting and cleansing the data and should be prepared to consult an independent resource or reviewer. This reviewer should be able to assess the state of the data independently, without the concern of prejudice or preference.

It also is important to note steps in the data analysis process during which errors or accidental changes to the data may occur. Any time that data is collected by means of questionnaires (paper and pencil or electronic), observations, interviews, etc; errors are likely to occur when data are entered into a database following collection, particularly if this process is done by hand. Data that are automatically entered into a database, if the collection method involves a computerized or online technique, may be less prone to errors at this step. Data will often need to be reformatted or pertinent variables selected from a larger dataset before it can be analyzed using a statistical program. This introduces another step in the data cleansing process and care should be taken to avoid altering data.

When the evaluation team is satisfied that the data have been adequately inspected and aberrant or missing values have been addressed, statistical analysis of the data can begin. Generally, analysis of data entails a two-step process, moving from the least complicated to the most complicated techniques. Initial statistical analyses, called descriptive statistics, involve describing and summarizing the data. The second step, inferential analysis, uses tests to compare and interpret the data. More information about each of these steps is provided below.

Descriptive Statistics

Descriptive statistics include a series of techniques that provide a summary or a “big picture” view of the data. Descriptives are a useful first step in data analysis because they can identify overarching themes or trends in the data that may help to guide later analyses. Generally data from program analyses fall into one of two data categories: categorical or continuous data. Categorical data generally can be divided into discrete categories with finite responses. Continuous data refer to data that fall on a continuous spectrum with an infinite number of possible responses or data points. Data can be broken down further into one of four levels of measurement: nominal, ordinal, interval and ratio. Data within each of these four levels of measurement share common characteristics as discussed in **Figure 26. Levels of Data Measurement**. The type of data will inform what descriptive and inferential statistics should be

conducted. Therefore, it is imperative that levels of measurement are identified early in the analytical process.

Figure 26. Levels of Data Measurement

Level of Measurement	Description of Data	Examples
Nominal	<ul style="list-style-type: none"> Data are discrete and generally refer to categories Data can be classified but not analyzed using mathematical operations 	<ul style="list-style-type: none"> Gender Race or ethnicity Patient category Service Branch
Ordinal	<ul style="list-style-type: none"> Data are discrete and generally refer to categories Data can be classified and ranked or ordered Generally data cannot be analyzed using mathematical operations Often numerical 	<ul style="list-style-type: none"> Military rank Level of education
Interval	<ul style="list-style-type: none"> Data can be classified and ranked or ordered Meaningful differences between values Always numerical 	<ul style="list-style-type: none"> Blood pressure readings Time of day Date of assessment
Ratio	<ul style="list-style-type: none"> Data can be classified and ranked or ordered Meaningful differences between data with a zero value Ratios can be calculated between data points Always numerical 	<ul style="list-style-type: none"> Height Weight

Within descriptive statistics, there are three major categories of analyses: measures of frequency, measures of central tendency and measures of variability.

Frequency refers to the number of cases or values within a category. Once an initial count of the frequency for each case or value has been determined, it is possible to calculate the percentage of cases in each category by dividing the number of cases in each category by the total number of cases (i.e., ratio). Outcomes of frequency analyses are often displayed in either tables, bar graphs/histograms or pie charts. **Figure 27. Frequency Table for Patient Category Data, Figure 28. Bar Chart for Patient Category Data, and**

Figure 29. Pie Chart for Patient Category Data present the same data (patient category data for a treatment evaluation) in each of the three formats to allow for comparison of the display options.

Figure 27. Frequency Table for Patient Category Data

Patient Category	Count
Active Duty	156
FAM MBR of Active Duty	68
Retired	23
FAM MBR of Retired	15
Other	5
TOTAL COUNT	267

Figure 28. Bar Chart for Patient Category Data

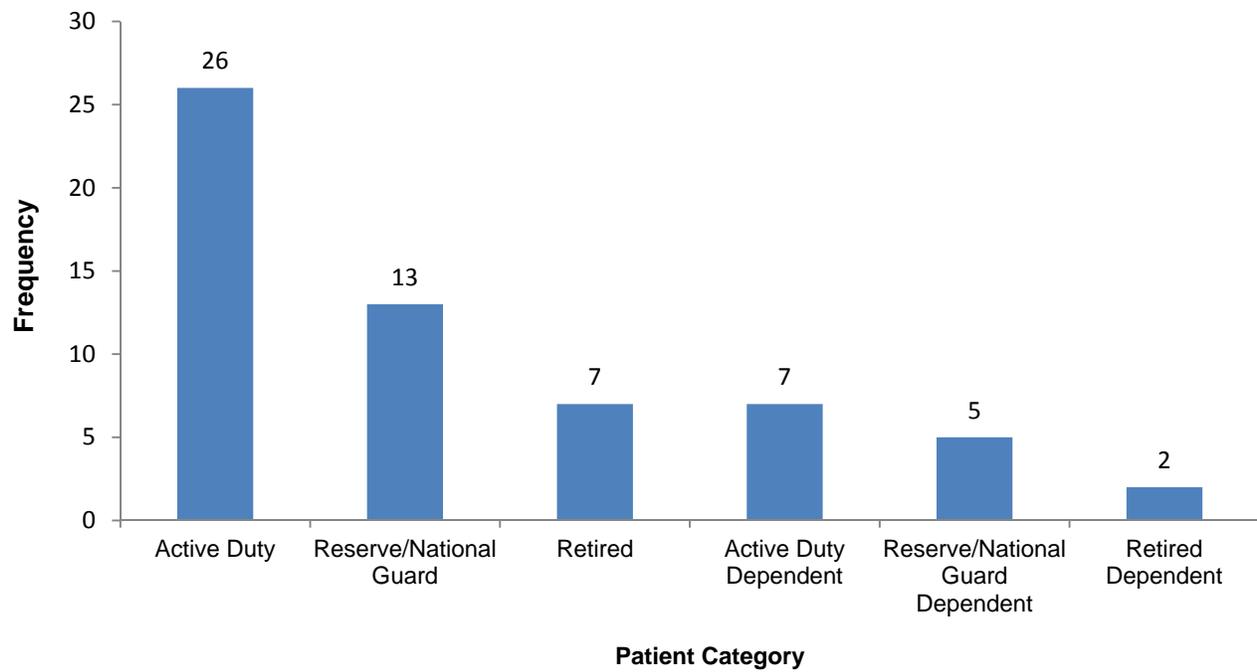
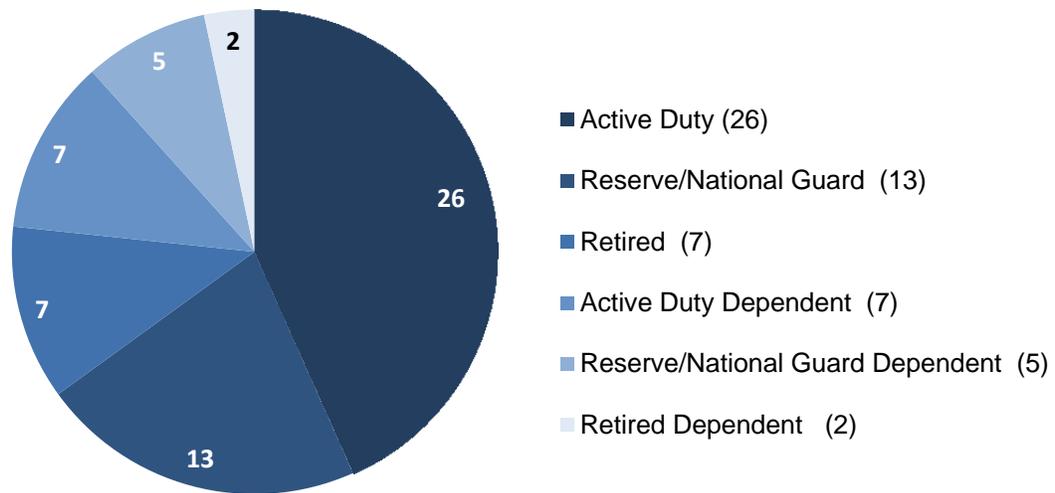


Figure 29. Pie Chart for Patient Category Data



Measures of central tendency calculate a single value that attempts to describe a set of data by identifying the central position within that data set. The three most common measures of central tendency are the mode, median, and mean scores. The correct measure of central tendency for a particular set of data will be determined by the level of measurement of the data. The *mode* is the score that occurs most frequently in a data set and can be determined regardless of the level of measurement. The *median* is the score the separates the upper half of the scores from the lower half. The median requires ranking of data before the middle score can be identified; thus it can only be used with ordinal, interval, or ratio data. Finally, the *mean* is the average of the scores and is calculated by summing the values and dividing by the total number of scores. The mean can only be calculated for interval or ratio data given the mathematical calculations necessary to derive it. The box at right provides examples of each.

EXAMPLE: Measures of Central Tendency

Dataset: The following data represent the number of appointments, by month, in a clinic over the course of the first five months of 2011:

JAN	FEB	MARCH	APR	MAY
35	45	35	30	20

MODE: The value 35 occurs twice, making it the mode of this dataset.

MEDIAN: When the data are ranked from low to high, 20, 30, **35**, 35, 45, the middle value is 35. The median for the dataset is 35.

MEAN: When the data are summed ($20+30+35+35+45 = 165$) and divided by the total number of scores ($165/5 = 33$), the calculated mean is 33.

Generally, a measure of central tendency does not provide enough information to sufficiently describe or summarize a data set. To present a more complete picture, a measure of central tendency is often accompanied by a measure of variability. *Measures of variability* describe how widely a set of measurements varies around a particular measure of central tendency, most often the mean. Measures of variability are generally only applicable to interval and ratio data. The most simplistic measure of variability is the *range* which is calculated by subtracting the minimum score from the maximum score of a dataset. More commonly, variability may be

expressed in terms of the *standard deviation from the mean* or more simply described as the *standard deviation*. The variability of the data, whether they are tightly clustered around the mean or more spread out, can provide important information about the overall nature of the data set and may help to inform which additional statistics will be used.

Calculating descriptive statistics for a data set provides the information needed to prepare figures and tables that represent and illustrate themes within the data. These statistics, particularly measures of central tendency and variability, can be translated into graphs and tables that provide an overview of the data and a foundation for the inferential statistical tests, the next step in the analytical process.

Inferential Statistics: Examining Relationships

Unlike the relatively simple goal of descriptive statistics to summarize the sample data, inferential statistics provide a means to interpret the data and draw conclusions that might be generalizable to larger populations. With inferential statistics, evaluators can look for differences between groups and make judgments regarding the probability that an observed difference between groups is a dependable one rather than one that occurred by chance. To choose which inferential statistic is most appropriate for a particular dataset, evaluators must consider several factors or characteristics of the data. These factors include:

- The study design
- The number of groups to be compared or related in the analysis
- The type or level of data being analyzed (nominal, ordinal, interval or ratio)

Inferential Statistics: Test Examples

The following section describes some of the more common inferential statistical tests, including the criteria that must be met for each. Please note that this is not an exhaustive list of tests but is meant to serve as an initial resource for determining which test or tests should be used for the analysis of program evaluation data. For additional information, more complex data sets, or data sets that violate assumptions for the tests described below, a trained statistician or analyst should be consulted.

Analysis of Variance (ANOVA): The analysis of variance tests for differences between multiple (two or more) groups, variables, or time points. There are several types of ANOVAs including univariate (one-way) ANOVA, multivariate ANOVA (MANOVA), repeated measures ANOVA, and analysis of covariance (ANCOVA). The exact nature of the data will determine which ANOVA is appropriate for each dataset. Generally, data should be interval or ratio data to be analyzed using an ANOVA.

Chi-Square: The chi-square test is a test of proportions and probability.

Correlation: Also referred to as the Pearson Product Moment Correlation, a correlation determines the direction (positive or negative) and magnitude of the relationship between two variables.

Regression: There are several types of regression analyses including simple regression and multiple regressions. Regression analyses examine the relationship between multiple variables, usually interval or ratio data. Regressions provide information about the effects of changing one variable when the others are held constant.

t-test: t-tests are used to test for differences between two groups, variables, or time points. Depending on the specific characteristics of the data, a single-sample, repeated measures, or independent group t-test may be required.

Figure 30. Inferential Statistics Criteria presents criteria that should be met for each of the inferential tests described above as well as program evaluation-relevant examples for each test.

Figure 30. Inferential Statistics Criteria

Inferential Statistic	Data Criteria	Examples
Analysis of Variance (ANOVA)	<ul style="list-style-type: none"> Dependent variables are continuous Independent variables or predictors are categorical Two or more groups or time points are being compared Specific study design will determine which ANOVA test to use 	<ul style="list-style-type: none"> Evaluators want to compare scores on a patient outcome measure for three different time points (pre-, during, and post-treatment); they would use a repeated measures ANOVA Evaluators want to compare scores on a patient outcome measure for three different patient groups (control, treatment 1, and treatment 2); they would use a univariate ANOVA
Chi-square	<ul style="list-style-type: none"> Dependent variables are categorical Independent variables or predictors are categorical Evaluators are interested in the frequency/proportion of a particular value or response 	<ul style="list-style-type: none"> Evaluators want to determine if the prevalence of patients being seen for depression in their satellite clinic is equal to the prevalence of patients with depression being treated in a larger military treatment facility
Correlation	<ul style="list-style-type: none"> Dependent variables are continuous Independent variables or predictors are continuous Evaluators are interested in the direction and magnitude of the relationship between two variables 	<ul style="list-style-type: none"> Evaluators want to determine the nature of the relationship between years of service and participant satisfaction with their program
Regression	<ul style="list-style-type: none"> Dependent variables are continuous Independent variables or predictors are continuous Evaluators are interested in the relationship between a dependent variable and multiple independent variables Specific study design will determine which type of regression to use 	<ul style="list-style-type: none"> Evaluators want to predict scores on a post-treatment outcome measure based on pre-treatment scores on the same measure; would use a simple regression model Evaluators want to know how age, rank, and pre-treatment scores relate to scores on a post-treatment outcome measure; would use a multiple regression model

Inferential Statistic	Data Criteria	Examples
t-test	<ul style="list-style-type: none"> • Dependent variables are continuous • Independent variables or predictors are categorical • Two or groups or time points are being compared • Specific study design will determine which t-test to use 	<ul style="list-style-type: none"> • Evaluators want to compare scores on a patient outcome measure for a single group of patients at two different time points (pre-, during, and post-treatment); use a repeated measures (within subjects) t-test • Evaluators want to compare scores on a patient outcome measure for two different patient groups (control and treatment); use an independent samples (between subjects) t-test

Statistical Software Packages

There are many options for statistical software packages that will perform basic and advanced statistical analyses. One of the most accessible options is *Microsoft Excel*, the spreadsheet software available in Microsoft Office. With minimal training, evaluators can use *Excel* to calculate a variety of both descriptive and inferential statistics, including most of the statistics described in this section. Use of *Excel* for statistical analyses does require some knowledge of statistical analysis in general and specific structures for entering data into the *Excel* spreadsheet to facilitate each test. This may make it less ideal for more complicated analyses.

Statistical software packages offer more options including more “user-friendly” interfaces, a larger number of functions and options, and the ability to write instructions and code to customize analyses based on individual data sets. Some of the more common statistical packages in the social science and medical fields are MatLab, R, SAS, Statistical Package for the Social Sciences, Stata, and Statistica. Some packages are open source and available without fee while others require permission and license fees. Evaluators should research the properties and capabilities of candidate software packages and consider factors such as accessibility and cost before choosing a package.

Interpreting Results and Drawing Conclusions

The ultimate purpose of the statistical analysis is to produce results that address the evaluation questions formulated earlier in the program evaluation process. Once the appropriate statistical tests have been conducted, evaluators should have the output necessary to answer or evaluate their program evaluation questions. Depending on the nature of the question, the data collected to address the question, and the statistics used to analyze the data, it may be a relatively simple answer or it may involve a greater level of interpretation.

The various types of designs are perhaps the greatest determinant of the level of confidence in ones findings. Descriptive studies examine the programs inputs, activities and process metrics that can inform evaluators about how well a program is functioning, but cannot speak to the effects of a program, as this type of evaluation does not address outcomes.

Experimental designs can address how effective a program is in meeting its objectives, as these designs incorporate outcome data. Due to the lack of random assignment of subjects in quasi-experimental designs, they are unable to conclude that the effects of a program are due to the treatment interventions. True experimental designs, which include random assignment of participants into groups, provide the strongest level of assurance that the programs outcomes are in fact due to the activities of the program.

Interpretation of the data analysis should not happen in a vacuum. Evaluators should strive to take the overall context of the program into account when interpreting the results of data analysis. There are many situations that can limit the degree of certainty an evaluator has in the results, even if they are statistically significant. A few such situations are illustrated below.

- **Baseline is not available:** In cases where the baseline of the measures in question is unknown, it is impossible to prove the program was responsible for the change. Consider a prevention campaign for substance abuse which aims to decrease binge drinking on a large military base. The four month campaign used public service announcements, on base classes and stand downs, as well as incentive programs to reduce drinking. Reported episodes of binge drinking on base steadily declined for each of the four months that the program was in operation. While it is tempting to conclude that the program was therefore a success, it is of course unknown whether the rate of binge drinking on base was already trending down. If the rate was already declining, it is much more difficult to prove that the program was effective.
- **Control group is not available:** While program evaluations do not have to have a control group or comparison group, the findings are much more credibly attributed to the interventions if there is a no-treatment control group. Without a control group, one cannot definitively know that the program was responsible for the results. For example, suppose that a hospital starts a new treatment protocol for patients who have recently sustained a mild TBI (mTBI). The program has treated 50 patients with mTBI, all of whom showed improvement over baseline scores on memory and attention, however, there is no comparison group available. Especially with the natural recovery rate for mTBI being so high, one cannot definitively say that the program's intervention was the cause of the improvement. The evaluation's findings are useful, as the findings can serve as a basis for generating interest and/or funding for a larger evaluation that includes a control group.
- **Program operates alongside several other interventions that might reasonably affect outcomes:** This is commonplace in non-academic centers, and presents a challenge to interpretation. Consider an evaluation of a treatment program for PTSD, which showed significant declines in PTSD symptomatology and higher return-to-duty

rates. This program was created at the same time as a treatment program for mTBI within the same military treatment facility. Approximately 85 percent of the patients with PTSD also had mTBI, and most of these patients were also receiving care for their mTBI concurrently. For the PTSD program, which showed significant positive effects, how can they be assured that the benefits are due to their program and not the mTBI program? The evaluators can account for patients who were also enrolled in the new mTBI program by analyzing subgroups (test to see if the beneficial effects of the PTSD program hold for those who are not also enrolled in the mTBI program) or statistically covarying the status of mTBI clinic enrollment. Entering mTBI program enrollment status as a covariate in an ANOVA will allow for separating out the variance from this intervention.

When an analysis fails to demonstrate the anticipated level of beneficial effects for a particular program, there are several possible interpretations that the program evaluation team should consider. For example, if a PTSD treatment program failed to show the expected levels of improvement in symptom levels or return to duty rate, there would be several possible reasons for this failure to demonstrate effectiveness. Some potential reasons are:

- **Lack of adherence to best practices in the program's interventions:** For example, if a large percentage of the programs providers are not following the program's treatment plans or are not trained in or using best practices / evidence-supported therapies.
- **The population is heterogeneous with respect to some key factor:** For example, if many of the patients are involved in the medical disability process, there will be an expected decline in prognosis and return-to-duty rates.
- **There may have been differences in "dosage" for the interventions:** For example, in most programs, there are many patients who do not attend all sessions and/or do not do all homework assignments. Were the patients who attended and participated more successful? **Figure 31. Participation, Outcome, and Effectiveness** below shows how variations in participation and outcomes can lead to different interpretations in data.

Figure 31. Participation, Outcome, and Effectiveness

If the participation was:	And the outcome was:	One might interpret this as:
Low attendance and or participation	Good outcomes	The intervention has very strong effect size OR some other factor is responsible
Low attendance and or participation	Poor outcomes	The program MAY be effective-if participants can get enough treatment
High attendance and or participation	Poor outcomes	The program not likely effective for this target group

- **There may be a difference in the type of patients seen within the program:** For example, if a large medical center has a lower return to duty rate than surrounding smaller military treatment facilities, one might be tempted to conclude that the medical center's treatment program is less effective. In this case, a direct comparison based on the rate of service members who are returned to duty may not be appropriate. Large medical centers routinely receive the most severe cases from local commands, whose medical departments cannot address the acuity of this set of patients' needs. The medical centers would therefore be expected to have a much higher rate of patients being medically boarded out of the service compared to its smaller neighbors.

Conclusion

Step 6: Analyze Data provides the information required to answer the original questions developed during the Preparation Phase. There are many methods of analysis, but the key is to use the method which provides the output necessary to answer the evaluation questions and best fits the criteria of the evaluation and the data collected. Synthesizing the results of the analysis into useful information will lead the evaluation team into the final step of the process and provide input for feedback and reporting.

Key Takeaways

- Prepare the data by eliminating missing values, redundant data and outliers
- Determine the appropriate data analysis methods for the type of data collected
- Synthesize and summarize the analysis to interpret the results

Sources and suggested readings

National Institute on Drug Abuse. (1993). *How good is your drug abuse treatment program?: Overview and case study*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy, Education, and Legislation, Community and Professional Education Branch.

Olney, P. C., & Barnes, M. S. (2008). *Collecting and analyzing evaluation data: Planning and evaluating health information outreach projects booklet 3*. Bethesda: National Library of Medicine, Outreach Evaluation Resource Center.

W.K. Kellogg Foundation. (1998). *Evaluation Handbook [Item #1203]*. Battle Creek, MI: Collateral Management Company.

Wolfe, B.L., & Miller, W. R. (n.d.). *Program evaluation: A do-it-yourself manual for substance abuse programs*. Albuquerque, NM: University of New Mexico, Department of Psychology.

Phase 3: Evaluation Feedback



It is at this point in the process where the evaluation team must gather all of the information collected during the Preparation and Execution phases to provide feedback to the program leaders and the interested stakeholders. **Phase 3, Evaluation Feedback** consists of two steps: **Step 7 - Develop Report**, and **Step 8 - Act on Findings**.



Step 7: Develop Report



After the team completes information gathering, result synthesis and preliminary result sharing with stakeholders, the team refines the results and drafts a report of the evaluation findings. This step requires the same careful planning approach used during the Preparation and Execution phases of the program evaluation.

Develop Report

- Draft a report of the evaluation findings
- Disseminate report for comments
- Finalize report and submit for approval

This step includes dissemination of the findings to the program leaders and stakeholder groups, and development of recommendations to act on the findings.

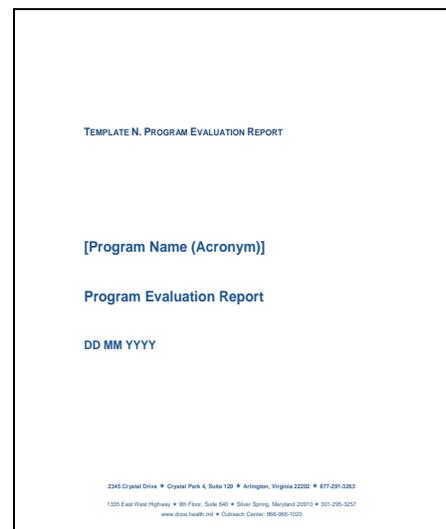
The **Program Evaluation Report [Template N]** is provided to document the program evaluation results. This report is a 10 to 15 page document that provides the stakeholders a detailed analysis of the program’s need, effectiveness and scalability. The report covers each of these program dimensions from the perspectives of strengths, lessons learned, opportunities and effectiveness measures. This report also addresses the initial program evaluation purpose and focus, as agreed upon with the program evaluation stakeholders during the Stakeholders Analysis.

Develop Report

The **Program Evaluation Report [Template N]** can be used to write the report. The template can also be adjusted to meet the requirements and needs of the group to whom the report is given. The following is a description of each of the sections in the template.

- **Executive Summary.** Program evaluation goals and focus, brief program description and summary of results
- **Program Evaluation Approach.** Brief description of interview results with program leaders and stakeholders, statement of the problem and brief description of methodology used to gather and analyze data
- **Program Summary.** Program goals and objectives, design, history and milestones, current status, populations served, policies and procedures and key partners
- **Program Evaluation Results.** Strengths, lessons learned, opportunities, changes to program approach, design, policies and procedures with documented impact and applicability to similar programs, review of effectiveness measures, data and associated program objectives, assessment of measure quality, quantitative

Template N: Program Evaluation Report
(double click on image to open)



analysis of program effectiveness and recommendations for measure refinement

- **Summary.** Synthesis of results (but not a restatement of the executive summary) with emphasis on impact and performance against program goals and institutional strategy
- **References.** Program documents, interviews, site visit information and secondary literature (avoid including the names of program staff without their permission)

To coordinate development of the draft Program Evaluation Report, the team lead should assign report sections to each team member, allocating sufficient time for writing and editing. Typically, the team lead and the SME write the technical sections of the report and the analyst writes the program background section, executive summary and concluding summary. Once the team members draft the report sections, the analyst combines the sections and reviews the entire report for style, consistency and tone. The analyst is also responsible for report design and formatting. Following the analyst’s review, the team lead carefully reviews and edits the report for technical insight, thoroughness, objectivity, accuracy and overall quality. Within Defense Department programs, this report will be finalized by the program director/manager.

Tips for writing an effective evaluation report include: (Kellogg, 1998)

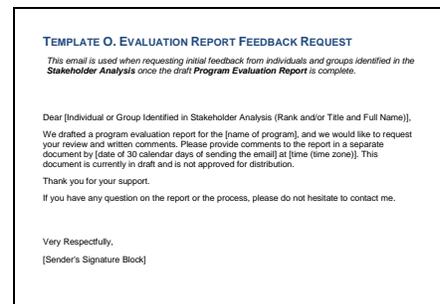
- Know who the audience is and what information is needed. Scope the report to meet the expectations of the intended audience. This may be the program leader or a headquarters level office chartering the evaluation. Refer back to the information gathered during the Review of Program step to ensure all expectations are addressed.
- Start with the most important information in the Executive Summary. While writing, imagine that the audience will not have time to read the whole report; be brief, yet informative. Develop concise reports by writing a clear Executive Summary, then start each chapter, subsection, or paragraph with the most important point.
- Make the report easy to read. Do not use professional jargon or vocabulary that may be difficult to understand. Use active verbs to shorten sentences and increase their impact. Write short paragraphs, each covering only a single idea.
- Edit the report, looking for unnecessary words and phrases. It is better to have someone else edit the work; however, if the originator must edit it themselves, it is best to allow a day or two to pass between writing and editing.

Disseminate Draft Report for Stakeholder Feedback

Upon completion of a draft of the **Program Evaluation Report [Template N]**, the team lead sends the report draft along with a cover letter **Evaluation Report Feedback Request [Template O]** to the individuals and groups identified in the **Stakeholder Analysis [Template E]**, which was developed during Step 1, Review of Program. Prior to sending this draft report out to stakeholders, the team lead should secure chain of command approval to forward this information.

DCoE recommends providing 30 days for stakeholders to review and provide written comments. Program feedback

Template O: Evaluation Report Feedback Request (double click on image to open)



may range from positive support and agreement with the program evaluation to minor corrections to direct disagreement with sections of the report. The team then reviews submitted comments and discusses whether it is necessary to make any changes to the report. The team's priority is to maintain the report's integrity, but the team should address comments from program stakeholders even if no changes are made. The team lead decides which edits to incorporate and which written comments to attach to the report. The team typically spends no more than one week making report revisions in order to continue the report finalization process in a timely manner.

To finalize the **Program Evaluation Report [Template N]** for dissemination, the team must obtain approval from appropriate stakeholders; such as senior leaders within the chain of command, public affairs, legal, and in cases of external evaluations, the primary program leaders. The analyst refers back to the **Stakeholder Analysis [Template E]** to identify stakeholders whose written approval is necessary to finalize the report. Likewise, the team lead and the analyst identify the necessary non-stakeholder approvals, such as those from agency-level strategic communication offices. The team should follow local policies regarding obtaining approvals, and then forward the Program Evaluation Report up the chain of command for approval, using existing routing forms.

Template P: Evaluation Report Approval Tracking Form
(double click on image to open)

TEMPLATE P. EVALUATION REPORT APPROVAL TRACKING FORM
The purpose of this worksheet is to assist program evaluators in documenting and tracking the approvals needed to finalize the Program Evaluation Report.

APPROVER NAME, OFFICE	LAST CONTACT	STATUS
LEADERSHIP		
Name, Office		
STRATEGIC COMMUNICATIONS		
Name, Office		
STAKEHOLDERS		
Name, Office		
OTHERS		
Name, Office		

For programs that are not under Defense Department or do not have local document approval routing procedures, a form has been provided in **Template P: Evaluation Report Approval Tracking Form**, to help monitor and track the approval process. The analyst can populate the local routing forms or the Evaluation Report Approval Tracking Form and, under the direction of the team lead, identify proposed turnaround times for each party's approval. The team lead monitors progress of the approvals and follows up as necessary. The team lead also addresses any feedback from requisite approvers. Once approvals are complete, the **Program Evaluation Report [Template N]** is officially final and ready for public dissemination.

Once the report is finalized, the team lead sends the completed report to the interested parties identified in the **Stakeholder Analysis [Template E]**. A sample cover letter is provided in **Template Q: Final Evaluation Report Distribution Letter**, for this communication. Dissemination of the report to all stakeholder groups emphasizes the collaborative nature of the evaluation methodology and encourages stakeholders to maintain ongoing dialogue about the program. Lastly, in order to facilitate DCoE's mission to identify, validate and disseminate effective program practices, the team should forward a copy of the final report to DCoE. Program Evaluation Reports from across the MHS and VHA received by DCoE will be used to identify and monitor promising programs for possible wider dissemination of best practices.

Template Q: Final Evaluation Report Distribution Letter
(double click on image to open)

TEMPLATE Q. FINAL EVALUATION REPORT DISTRIBUTION LETTER
This email is used to distribute the final Program Evaluation Report to interested parties identified in the Stakeholder Analysis after obtaining approval from appropriate individuals and offices identified in the Report Approvals Worksheet.

Program Evaluation [DD MM YYYY]

ProgramName
Final Report Communication

Dear [Individual or Group Identified in Stakeholder Analysis (Rank and/or Title and Full Name)],

Attached is the final program evaluation report for the [name of program] program. This report has been reviewed and approved by [noteworthy approvers and strategic communications officers], and can be distributed to [level of distribution allowed (publicly, privately among certain stakeholders)].

We want to thank you again for support provided throughout this program evaluation process. There is no further action required on your part. We hope you will keep us updated on any activities related to [name of program], as well as any noteworthy use or dissemination of the evaluation report.

If you have any questions about the evaluation report or process, please do not hesitate to contact me.

Very Respectfully,
[Sender's Signature Block]

Key Takeaways

- After obtaining approvals to send the draft report to stakeholders, request input and address comments on the **Program Evaluation Report [Template N]**
- Finalize the Program Evaluation Report and submit for the necessary approvals

Sources and suggested readings

Aubel, J. (1999). *Participatory program evaluation manual*. Calverton: Child Survival Technical Support Project and Catholic Relief Services.

National Institute on Drug Abuse. (1993). *How good is your drug abuse treatment program?: Overview and case study*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Drug Abuse, Office of Science Policy, Education, and Legislation, Community and Professional Education Branch.

W.K. Kellogg Foundation. (1998). *Evaluation Handbook [Item #1203]*. Battle Creek, MI: Collateral Management Company.

Step 8: Act on Findings



Lastly, the evaluation team should work with the program leaders to determine how to act on the findings and recommendations of the evaluation. Many program evaluations have stopped at the report dissemination step, leaving the program leaders and stakeholders to make the connection between the findings and actionable improvement plans.

A report that sits on a shelf after it is delivered will not lead to changes or improvements. An effective program evaluation is one that will lead to and support follow-on actions. The recommendations in the report should provide program staff and stakeholders the direct impetus to make actionable decisions about the program. The program evaluation program results will identify strengths and weaknesses, provide information on key factors affecting the program, and clarify options for moving forward. It is crucial that the program evaluation team and the program leaders begin discussions at this point on utilizing the recommendations outlined in the report. Specific uses of the evaluation findings will depend on the original purpose of the evaluation and the questions addressed.

Act on Findings

- Make actionable decisions about the program
- Improve the program and identify options for moving forward
- Evaluation effectiveness of the Program
- Generating new knowledge

Brainstorming

One of the most valuable tasks in this stage is to brainstorm with program leaders and stakeholders in order to draft implementation strategies for the evaluation results using the **Results Implementation Worksheet [Template R]**. This worksheet helps the evaluation team brainstorm, prioritize and plan follow-up actions stemming from the results detailed in the **Program Evaluation Report [Template N]**. The template is not a detailed process improvement plan but rather a structured means to build on the momentum of the program evaluation to facilitate program improvements. Program leaders, staff and stakeholders are more likely to use evaluation results if they have ownership over the evolution of the program as a result of the evaluation process. The more people who are actively involved in determining the future direction of the program, the easier it will be to facilitate using the results for program improvement and decision making.

The results of the evaluation may validate that the program is working well. In this case it is important to codify a plan to continue regular, routine data collection and analysis to ensure the program continues to meet its goals and objectives. If the evaluation discovered a problem or point of failure, the leadership team should explore this further to determine the cause and develop solutions. There are many methods that can be used to accomplish this, such as root cause analysis.

In any case, in order to act on the evaluation results, begin discussions with project leaders and stakeholders by asking questions such as:

- What do key stakeholders need to know more about based on the evaluation?
- What additional information is needed before making decisions?
- What needs to happen based on the answers to the evaluation questions? (Play out different scenarios, depending on the different answers received.)

Once these questions are answered, the evaluation team can articulate how to utilize the evaluation results by developing actionable steps using the **Results Implementation Worksheet [Template R]**. The worksheet lays the groundwork for the design and methodology of an implementation plan, specifies the timeline and metrics..

Template R Template R: Results Implementation Worksheet
(double click on image to open)

TEMPLATE R. RESULTS IMPLEMENTATION WORKSHEET					
Evaluation Findings	What type of action is needed?	What specific steps are required?	Assign action officers	Metrics	Timeline for implementation
Patients in the program are not showing the expected levels of clinical improvement.	After review of the program and evaluation results, it was determined that care needed to more closely follow evidence based guidelines.	<ol style="list-style-type: none"> 1. Perform a chart review to determine existing patterns of care 2. Conduct refresher training on VA/DoD CPOs 3. Add compliance with clinical practice guidelines to monthly chart review checklist 4. Verify change in patterns of care over time 	<ul style="list-style-type: none"> • LT Jones • Department Head • Clinic quality assurance lead • Department Head 	<ul style="list-style-type: none"> • % of cases not meeting standards of care • # of trainings / % of staff who attended • % of chart reviews using new form • % of chart reviews failed 	<ul style="list-style-type: none"> • one month • one month • one week • six months
1.					
2.					
3.					

If any program changes are made as a result of the evaluation, then key program components such as goals, objectives and effectiveness measures should be mapped to the institutional vision, mission and goals to determine whether the program remains in alignment with organizational priorities. Conducting periodic reviews on a quarterly basis (or more frequently as needed) will help program leaders determine whether elements of the program remain relevant given organizational changes and emerging or changing priorities.

Conclusion

By writing a thorough evaluation report and conducting follow-up discussions for implementing actionable recommendations, the evaluation team will have completed a careful process that brings the evaluation full cycle. Regular and continuous review of the program will ensure the best use of resources to benefit the program and its stakeholders.

Key Takeaways

- Initiate recommendations to act on findings through brainstorming using the **Results Implementation Worksheet [Template R]**
- Reviews of the program should be conducted on a regular basis and whenever there has been a significant shift in organizational priorities or needs

Sources and suggested readings

Wolfe, B.L., & Miller, W. R. (n.d.). *Program evaluation: A do-it-yourself manual for substance abuse programs*. Albuquerque, NM: University of New Mexico, Department of Psychology.

CONCLUSION

The Defense Department has invested significant resources into the creation and expansion of multiple programs across the MHS in an effort to improve PH and TBI care for service members and their families. The ability to demonstrate the effectiveness of these programs serves as a useful feedback mechanism when programs are competing for support and funding in a resource-constrained environment. Additionally, when clinical and statistical significance has been demonstrated, a program may become a model for the military health enterprise, reducing redundancy, increasing cost effectiveness and providing the best care available. The Program Evaluation Guide assists program leaders with developing or refining goals, establishing effectiveness measures and synthesizing data in a progressive feedback loop, which provides checks and balances to determine whether a program is effectively meeting its stated objectives.

APPENDICES

APPENDIX I: EXAMPLES OF PROGRAM EVALUATION (PE)

Program / Organization	Type of PE	Purpose / Methodology	Findings	Reference
US Air Force Suicide Prevention Program (AFSPP)	Impact	To determine AFSPP's impact on suicide rates in the Air Force. Applied an intervention regression model to data from 1981 through 2008. This provided 16 years of data prior to program launch and 11 years post launch.	With the exception of 2004 (when the program was less rigorous), suicide rates in the Air Force were significantly lower after the program was launched than beforehand	Kerry L. Knox, P., Pflanz, S. M., Talcott, G. W., Campise, R. L., Lavigne, J. E., Bajorska, A. M., et al. (2010, December). The US Air Force suicide prevention program: implications for public health policy. <i>American Journal of Public Health</i> , 100(12), 2457-2463.
Mental Health Department of the U.S. Naval Hospital in Okinawa, Japan Outcome Crisis Prevention Program	Outcome	To determine whether the program activities reduced symptoms and promoted coping ability. Administered the Beck Depression Inventory II and the Beck Anxiety Inventory pre-training and post-training to patients who completed the program between February 2003 and February 2005 (N = 326)	The Outpatient Crisis Prevention Program effectively reduced depression and anxiety symptoms and promoted healthy coping behavior among participants. Follow-up revealed that this was maintained 1 month following treatment.	David E. Jones, C. M., Perkins, K. H., Cook, J. L., & Ong, A. L. (2008). Intensive coping skills training to reduce anxiety and depression for forward-deployed troops. <i>Military Medicine</i> , 173(3:241).
Substance Abuse Rehabilitation Program, U.S. Naval Hospital, Okinawa Japan	Process	To determine whether the program was meeting its mission to be effective in preventing suicides in both military members and eligible beneficiaries and effective in facilitating the retention of military members with gambling problems. Reported debt and losses, frequency of suicidal behavior, depressive disorder, substance abuse, and scores on South Oaks Gambling Screen (SOGS), Alcohol Use Disorders Identification Test, Therapy Efficacy Monitoring Scale (TEMS), and Beck Depression Inventory II (BDI-II) were examined in gambling treatment participants.	A significant degree of depression, suicidality, and substance abuse problems were observed in the sample.	Carrie H. Kennedy, M. U., Cook, J. L., Poole, D. S., Brunson, C. A., & Jones, D. L. (2005). Review of the first year of an overseas military gambling treatment program. <i>Military Medicine</i> , 170(8:683).

Program / Organization	Type of PE	Purpose / Methodology	Findings	Reference
National Naval Medical Center (NNMC) historical data	Outcomes	To determine whether outcomes vary for patients who received care via video conferencing v. care provided in person. Historical data was reviewed. One group was seen via video conferencing and a second group was seen face to face (the control). Chi-square tests determined significant associations.	The Global Assessment of Functioning (GAF) was significantly more improved for the video conferencing group than for the face to face. Providers using video conferencing tended to tell more patients to return for follow-up appointments in 30 days or less, leading to improved compliance and shorter time to follow-up.	Grady, B. J., & Melcer, T. (2005). A retrospective evaluation of telemental healthcare services for remote military populations. <i>Telemedicine and e-health, 11</i> (5).
A community-based wellness program attended by Army soldiers and their spouses	Outcomes	To assess the stage of change progression and elimination of unhealthy behavior after attending the program. A quasi-experimental prospective study (N=245) was conducted and differences in stage progression between the intervention group and a comparison group were examined	Significant differences were seen in the fitness and communication risk categories immediately following the intervention and in fitness and spirituality 6 months after baseline. There was a significant increase in seat belt use and decrease in tobacco use and a reduction in family stress and personal stress in the intervention group. The pilot study revealed many positive non-significant trends in risk reduction.	Niederhauser, V. P., Maddock, J., LeDoux, F., & Arnold, M. (2005, March). Building strong and ready Army families: A multi reduction health promotion pilot study. <i>Military Medicine, 170</i> (3), 227-233.
Walter Reed Army Medical Center (WRAMC)	Outcomes	To assess the effectiveness of inpatient cognitive rehabilitation for patients with moderate and severe TBIs. A single-center, parallel-group randomized trial of 273 consecutively hospitalized TBI patients that were randomly assigned to an intensive, standardized 8-week in-hospital cognitive rehabilitation program (n=67) or a limited home rehabilitation program with weekly telephone support from a psychiatric nurse (n=53). Outcome measures were return to gainful employment and fitness for military duty at the 1-year follow-up.	At the 1 year follow-up, there was no significant difference between those in the intensive in-hospital cognitive rehabilitation program v. the limited home rehabilitation program I return to employment. Additionally, there were no significant differences in cognitive, behavioral or quality-of-life measures. A post-hoc analysis of patients who were unconscious > 1 hour post TBI revealed the in-hospital group had a greater return-to-duty rate (80% v. 58%, p=.05).	Salazar, A. M., Warden, D. I., Schwab, K., Spector, J., Braverman, S., Walter, J., et al. (2000, June 21). Cognitive rehabilitation for traumatic brain injury: A randomized trial. <i>American Medical Association, 283</i> (23).

Program / Organization	Type of PE	Purpose / Methodology	Findings	Reference
Restore & Renew Wellness Clinic (R&RWC) at a United States Department of Defense hospital.	Process	To evaluate the practicality of a weekly on-site complementary and alternative medicine (CAM) wellness clinic for staff at a military hospital, and to capture employees' perceptions of program effectiveness. During visits, participants (nurses, physicians, clinicians, support staff, administrators) selected modalities (ear acupuncture, clinical acupressure, and Zero Balancing). Following the visit, a self-report survey was done to evaluate clinic features and perceived impact on stress-related symptoms, compassion for patients, sleep, and workplace or personal relationships.	2,756 surveys revealed most participants agreed or strongly agreed to feeling more relaxed after sessions and would recommend it to a co-worker. Most participants strongly agreed to experiencing increased compassion with patients, better sleep, improved mood, and more ease in relations with co-workers. Perceived benefits were maintained and enhanced by the number of visits. The results suggest a CAM hospital-based wellness clinic is feasible, well-utilized, and perceived to have positive health benefits regarding stress reduction at work, improved mood and sleep, and lifestyle.	Duncan, A. D., Liechty, J. M., Miller, C., Chinoy, G., & Ricciardi, R. (2011). Employee use and perceived benefit of a complementary and alternative medicine wellness clinic at a major military hospital: Evaluation of pilot program. <i>The Journal of Alternative and Complementary Medicine</i> , 17(9), 809-815.
A volunteer victim advocacy program at a US Army Installation	Process	To assess VA programs for training needs, difficult situations, best services, and ability to meet personal needs. 87 volunteers contributed data and were monitored through initial, periodic, and exit interviews	Results reveal that it is feasible to develop a volunteer VA program to assist clients at the scene of a domestic violence incident.	McCarroll, J. E., Castro, S., Nelson, E. M., Fan, Z., Evans, P. K., & Rivera, A. (2008, September). Establishing and maintaining a volunteer victim advocate program to assist victims of domestic violence in the U.S. Army. <i>Military Medicine</i> , 173(9), 860-864.
U.S. Navy Sexual Assault Intervention Training (SAIT) program for women	Outcomes	To evaluate the program's effect on measures of rape knowledge, empathy for rape victims, and acceptance of rape myths. A randomized clinical trial evaluated the effect of the SAIT program. 550 female personnel who had completed basic training were divided into 2 groups: those who participated in the SAIT program and a comparison condition. Both groups completed measures of rape knowledge, empathy for rape victims, and acceptance of rape myths.	SAIT increased factual knowledge about rape, empathy with rape victims, but did not reduce women's rape myth acceptance. The findings suggest that the program is having an impact.	Rau, T. J., Merrill, L. L., McWhorter, S. K., Stander, V. A., Thomsen, C. J., Dyslin, C. W., et al. (2011). Evaluation of a sexual assault education /prevention program for female U.S. Navy personnel. <i>Military Medicine</i> , 176(10), 1178-1183.

Program / Organization	Type of PE	Purpose / Methodology	Findings	Reference
<p>The Strategies to Assist Navy Recruits' Success (STARS) project for the BOOT Camp Survival Training for Navy Recruits</p>	<p>Impact</p>	<p>To evaluate the effect of the BOOT STRAP mental health intervention on cohesion, problem-solving coping strategies, perceived social support, and anger expression coping strategies. A prospective cluster-randomized intervention trial of 1,199 recruits divided into two groups, those who received the intervention (A prescription "BOOT STRAP" mental health intervention) during surge training and those who did not.</p>	<p>The intervention group revealed significantly higher group cohesion and higher perceived social support, but reported lower scores on anger expression coping strategies. Potential cost savings were estimated at 18.6 million and the intervention would cost 1.5 million. The study shows program potential for decreasing attrition, improving recruit performance levels, and provides a cost-effective method of enhancing recruit retention.</p>	<p>Williams, R. A., Hagerty, B. M., Andrei, A.-C., Yousha, S. M., Hirth, R. A., & Hoyle, K. S. (2007, September). STARS: Strategies to assist Navy recruits' success. <i>Military Medicine</i>, 172(9), 942-949.</p>
<p>S.A.F.E. Program created in the Veterans Affairs (VA) system</p>	<p>Outcomes</p>	<p>To evaluate program changes and evaluation data including participant retention and satisfaction. 5 years of data for 170 family members was evaluated for awareness of VA resources, level of caregiver distress, understanding of mental illness, ability to engage in self-care, which was correlated with number of sessions attended.</p>	<p>The data revealed that the S.A.F.E. Program attendance is positively correlated with understanding of mental illness, awareness of VA resources, and ability to engage in self-care activities. It is inversely correlated with caregiver distress.</p>	<p>Sherman, M. D. (2006, April). Updates and five-year evaluation of the S.A.F.E. program: A family psychoeducational program for serious mental illness. <i>Community Mental Health Journal</i>, 42(2), 213-219.</p>

Program / Organization	Type of PE	Purpose / Methodology	Findings	Reference
The Polytrauma Network (PNS) Clinic within the Department of Veterans Affairs (VA)	Process	To evaluate the PNS clinic's ability to demonstrate the VA healthcare system's evolving effort to meet the clinical needs of this population and to fine tune the clinic process as it is a new program. 62 participants were screened through a 20 minute interview and a 20 minute testing period. Measures included the Neurobehavioral Symptom Inventory (NSI), the PTSD Checklist (PCL), Beck Depression and Anxiety Inventories, Neuropsychological Assessment Battery-Screening Module, Wisconsin Card Sorting Test, Trail-Making Test, Grooved Pegboard Test, Wechsler Processing Speed Index and Working Memory Index, Wechsler Test of Adult Reading, and Shipley-Hartford Vocabulary and Abstraction tests, speech-language and hearing evaluation, a visual evaluation, and a Patient Satisfaction Survey.	The population revealed a high prevalence of post-concussion symptoms, posttraumatic stress, poor cognitive performance, head and back pain, auditory and visual symptoms, and problems with dizziness or balance. An anonymous patient feedback survey used to fine-tune the clinic process, reflected high satisfaction with this new program.	Department of Veterans Affairs. (2007). Program development and defining characteristics of returning military in VA Polytrauma Network Site. Journal of Rehabilitation Research & Development, 44(7), 1027-1034.
The Fort Lewis Soldier Wellness Assessment Program (SWAP)	Process	To evaluate the current SWAP process's value, whether it meets the DoD's annual health screening needs, and its ability to screen for additional areas of concern such as TBI. This evaluation also examined the extension of the program to two additional Army installations. 7,880 participants completed the SWAP Quality Improvement Questionnaire (QIQ), an anonymous 25 question soldier satisfaction questionnaire custom developed for SWAP.	85% of soldiers said they were honest on the HRAII v2, 85% were honest with their mental health counselor, 73% felt the survey was comprehensive, 81% felt the provider review was adequately detailed. 8% of soldiers felt the process was not helpful. 65% felt the completion of the process demonstrated the Army was interested in their health. 60% agreed they were more comfortable seeking health care in the future because of the experience. The SWAP process has been recognized and has been called to expand to additional sites.	Gahm, G. A., Swanson, R. D., Lucenko, B. A., & Reger, M. A. (2009, July). History and implementation of the Fort Lewis Soldier Wellness Assessment Program (SWAP). Military Medicine, 174, 721-727.

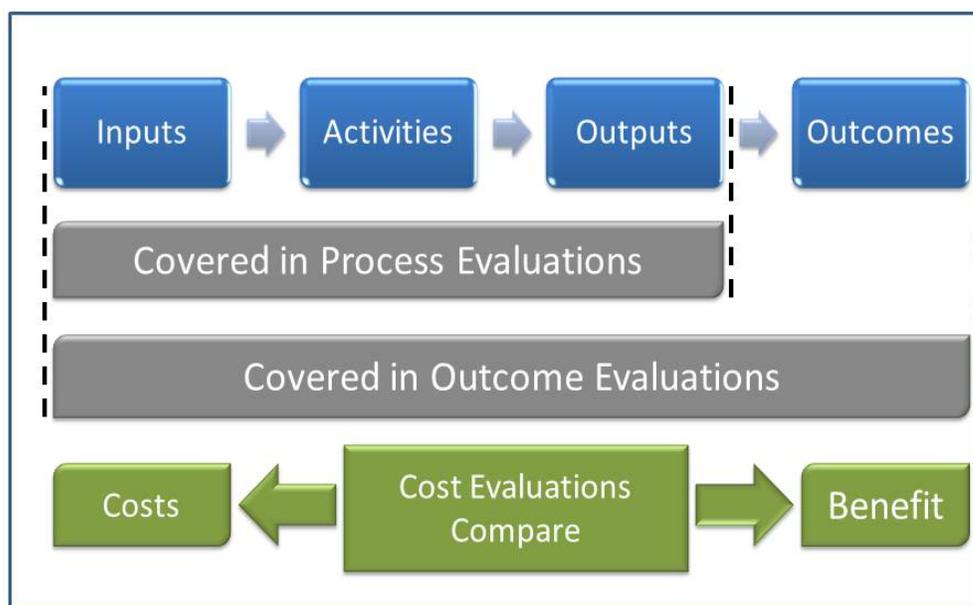
APPENDIX II: INTRODUCTION TO COST ANALYSIS FOR PROGRAM MANAGERS

Why Consider Costs in Healthcare?

Since funding for healthcare is subject to some limitations, it is ideal to identify programs that are able to achieve desired effects at the lowest costs. Understanding of program cost effectiveness will allow decision makers to concentrate funding on those programs which are producing meaningful outcomes at a low cost. Such action would maximize the number of people who can be served, particularly given limited resources.

While evaluations of newer programs usually focus on process, it is important to plan and begin data collection for cost analysis early. As a program matures, its activities should produce measurable outcomes that demonstrate that it is effective in achieving its objectives. Cost analysis examines a program in terms of how much benefit is produced for the expended funding. Because this level of analysis requires measures of program outcome, it is typically performed late in the lifecycle of a program, once cost and outcome data is available. Figure 32. **Cost Analysis Logic Model** demonstrates the aspects of a program that are analyzed in a cost analysis.

Figure 32. Cost Analysis Logic Model



This appendix provides a high level overview of cost analysis in healthcare decisions for program managers, specifically discussing two types of cost analysis: Cost-Effectiveness Analysis (CEA) and Cost-Benefit Analysis (CBA). The content was adapted from the Ceilini and Kee chapter on cost-effectiveness and cost-benefit analysis (2010). While this appendix reviews key concepts and explains the role of the program manager in the cost analysis, it is not an instruction guide. Information on how to conduct CBA and CEA and similar economic evaluation methods is beyond the scope of this Guide; however, there is a suggested reading list with references, including the Cellini and Kee chapter, at the end of this section for those who want to explore this topic in greater depth.

CEA and CBA are described briefly in the following sections as well as the role that an individual program manager plays in these types of analysis. This brief overview emphasizes the steps common to both CEA and CBA, noting the places where the two techniques differ as well.

What is Cost-Effectiveness Analysis?

CEA measures key outcomes or benefits of a program against the costs involved in producing these effects. In order to do this, the various inputs for a program must be quantified in terms of dollar value to compute a total cost for the program. The program's units of effectiveness are also specified, allowing evaluators to calculate a cost effectiveness ratio. Units of effectiveness provide a measure of quantifiable outcomes that are central to the program's goals and objectives. The CEA ratio is simply the total costs of the program divided by the units of effectiveness.

$$\text{Cost Effectiveness Ratio} = \frac{\text{Total Cost}}{\text{Units of Effectiveness}}$$

What is a Cost-Benefit Analysis?

A CBA is very similar to CEA in that they both assign a dollar value to the costs of a program. Unlike an analysis of cost-effectiveness, CBA takes the analysis a step further in that it also calculates a dollar value that represents the program benefits. A usual measure associated with cost benefit analysis is the net benefit of the program, calculated by the simple formula:

$$\text{Net Benefit} = \text{Total Benefits} - \text{Total Cost}$$

These equations may seem simple; however, the process to estimate program costs and benefits is complex. The calculations require assumptions about the value of individual benefits which can be quite subjective.

Conducting a CEA or CBA

Whether conducting a cost-effectiveness or cost-benefit analysis, common steps are required. These steps are listed in **Figure 33. Steps in Cost Analysis**

Figure 33. Steps in Cost Analysis

1. Set the analysis framework
2. Decide which costs and benefits should be recognized
3. Identify and categorize costs and benefits
4. When applicable, project costs and benefits over the life of the program
5. *Monetize* (establish a dollar value for) costs
6. Determine the units of effectiveness (for CEA), or monetize benefits (for CBA)
7. Discount costs and benefits to establish present values
8. Compute a cost-effectiveness ratio (for CEA), or determine the net present value (for CBA)
9. Perform a sensitivity analysis
10. Make a recommendation

Step 1: Set the analysis framework

Identify whether a CEA or a CBA is needed. The box to the right lists some characteristics of each type of analysis. These analyses can be performed before, during, or upon completion of a program. The appropriate framework will vary by program; however throughout any framework, the only costs and benefits that should be considered are those that occurred specifically as a result of the program which is under evaluation.

Cost Effectiveness vs. Cost Benefit

- **Cost Effectiveness** - Best suited for determining which of a set of programs achieves the best outcomes for the lowest cost. Very useful when it is difficult to establish a dollar value for the major outcomes.
- **Cost benefit** - Best suited for analyzing a single program to determine if the benefits of the program are worth the costs. If comparing multiple programs, cost benefit is helpful in determining which has the greatest benefit to society.

Step 2: Decide which costs and benefits should be recognized

Often, programs have a range of stakeholders whose definition of costs and benefits may vary. In determining the scope of an analysis, it is important to define what cost and benefit boundaries and definitions should be used. A more narrow scope will require fewer costs and benefits; however, a narrow scope also increases the risk that relevant costs and benefits may be missed.

Step 3: Identify and categorize costs and benefits

The next step in a CBA or CEA is to create a listing of every possible cost or benefit that falls within the boundaries identified in the previous step. Not all costs and benefits are equally important in terms of a cost analysis, and smaller costs and benefits can often be dropped from an analysis. Evaluators tend to focus on the costs and benefits that have significant implications to the program. During the early steps of the analysis, it is advisable to capture a list as comprehensive as possible that can later be reduced as needed.

Step 4: When applicable, project costs and benefits over the life of the program

This step requires consideration of the evaluation time frame, and potential changes in costs or benefits over time. Both the length and unit of time being analyzed can vary, although it is common to measure costs and benefits over the course of years in more mature programs. It may be necessary to predict impact over the life of a project and decide if the costs or benefits will increase, decrease or disappear in each subsequent unit of time. If a change in the costs or benefits is likely to occur, the next question is how this change will occur (i.e. at regular or irregular intervals).

Step 5: Monetize (establish a dollar value for) costs

Monetizing each identified cost, is relatively straightforward for some costs (i.e. investments, cost of capital, indirect cost allocation), but challenging for the more complicated costs (i.e. social costs). In order to have the most valid analysis, all costs must be expressed in the same units. When monetizing costs, it is important to clearly state the nature of the cost, how it is defined and measured, and any assumptions used in the monetizing procedure. A sensitivity analysis (described later) determines how the outcome of the analysis is impacted by the assumptions. This step is the same for both CBA and CEA.

Step 6: Monetize benefits

While establishing costs follows a similar process for CEA and CBA, the process for monetizing benefits differs depending on type of cost analysis being conducted.

- For a CEA, the most important benefit of the program should be identified and quantified using units of effectiveness. An example of this would be the number of diseases prevented or number of lives saved. Using common units of effectiveness allows for comparisons between programs that have a similar focus. For example, various programs providing substance use disorder treatment could be compared to one another by participant abstinence levels.
- For CBA, monetizing benefits follows the same strategy as monetizing costs. Each benefit must first be quantified and then assigned a dollar value. The goal of CBE is to calculate a dollar value for all major outcomes and benefits. A CBE becomes increasingly more difficult when a program has complex objectives impacting different populations.

Step 7: Discount costs and benefits to establish present values.

People tend to value costs and benefits in the present greater than the same costs and benefits that may be incurred in the future. A *social discount rate* is used to reflect the concept that society places a greater value on resources that are consumed or available today compared to resources that are available in the future. Applying the social discount rate to a monetary value calculates the present value. This present value should be used in the remaining CEA and CBA calculations. The specific *social discount rate* varies by organization. Identify whether this step is required by the organization and, if so, the correct rate to be applied.

Step 8: Compute a cost-effectiveness ratio or net present value.

In this step, one will either compute a cost-effectiveness ratio for CEA or a benefit-cost ratio for CBA. This links the present value of costs to benefits, as measured in dollars or in units of effectiveness. Cost effectiveness analyses use the following formula to compute a cost effectiveness ratio:

$$\text{Cost-effectiveness Ratio} = \frac{\text{Total Cost}}{\text{Units of Effectiveness}}$$

The results of a cost effectiveness ratio are often expressed as dollars per unit of effectiveness. Some examples include “dollars per overdose prevented” or “dollars per life saved”. Alternatively, the ratio can be reversed and simplified to express “overdoses prevented per dollar” or “lives saved per dollar.”

The benefit-cost ratio for the CBA is calculated by dividing the present value of the benefits by the present value of the costs using the following formula:

$$\text{Benefit-cost Ratio} = \frac{\text{Present value of the benefits}}{\text{Present value of the costs}}$$

When considering the benefit-cost ratio from an economic efficiency perspective, a ratio of greater than one would be considered an efficient use of resources.

It is important to note that cost-effectiveness ratios and benefit-cost ratios can mask differences in scale when comparing programs to each other. The ratios of two programs

could be the same while the costs and benefits from that program may be very different. Therefore, these ratios should be used only to compare programs of similar size.

Step 9: Perform a sensitivity analysis

Sensitivity analysis tests the sensitivity of a CEA or a CBE to particular assumptions. This step is critical to the interpretation of the analysis and provides insight into the effects of varying the assumptions of an evaluation. Assumptions are often “best guesses” and evaluators should be very careful to acknowledge the potential for error or changes to the analysis outcomes if certain assumptions are incorrect. There are two general kinds of sensitivity analyses, partial and extreme, as described in the call-out box to the right.

Partial vs. Extreme Sensitivity Analysis

- **Partial Sensitivity** - This analysis changes one number, parameter or assumption at a time, with all other aspects held constant. Best used when there are a limited number of critical assumptions.
- **Extreme Sensitivity** - This analysis changes all uncertain parameters simultaneously, using both best- and worst-case scenarios. Best used when there are greater levels of uncertainty.

Step 10: Make a Recommendation

The objective of a CEA or CBA is to make a recommendation or evaluation conclusion. If a program is found to have a positive net value (especially in a worst-case scenario extreme sensitivity analysis), then its merit is supported. Conversely, if a program has a negative net value (especially in a best-case scenario extreme sensitivity analysis), then its merit appears to be in question. There are several factors that should be considered when making a recommendation including the assumptions used in the evaluation, equity and distribution of costs and benefits, and the unquantifiable costs and benefits acknowledged.

What is the role of a program manager in cost analyses?

While the duties of most program managers will not include performing a CEA or CBA, it is important to understand the general principals involved and to know how these economic evaluations relate to the program.

Throughout the Guide, information on how to conduct both process and outcome evaluations has been presented, with the rationale for conducting such evaluations being:

- Verifying that the program is being implemented as planned
- Verifying that the program is meeting its goals
- Verifying that the program is effective (outcomes)

. Additionally, process and outcome evaluations as well as cost analysis provide information and tools that can help justify existing resources and request additional resources when needed.

Portfolio managers can use the information from these evaluations to prioritize programs. This is particularly relevant in times of economic restriction when the competition for program funds increases. Programs that can provide a greater depth and breadth of information about their processes, outcomes, and cost-effectiveness are more easily justifiable. A program that does not track information such as outcomes is at a disadvantage relative to programs that can provide this type of information.

Suggested Reading and References

Cellini, S. R. & Kee, J.E., (2010) Cost-effectiveness and cost-benefit analysis. In Wholey, J.S., Hatry, H. P., Newcomer, K.E. (Eds), *Handbook of practical program evaluation* (3rd ed., pp. 493-530). San Francisco, CA: Jossey-Bass.

APPENDIX III: OFFICIAL INSTRUCTIONS, POLICIES, REGULATIONS AND GUIDANCE RELATED TO PROGRAM EVALUATION

Organization and Title	Source
Program Evaluation	
<i>DoD Director of OSD Cost Assessment and Program Evaluation (CAPE)</i>	http://www.pae.osd.mil/
<i>Department of State Program Evaluation Policy (Bureau of Resource Management)</i>	http://www.state.gov/s/d/rm/rls/evaluation/2012/184556.htm
<i>Army Regulation 73-1: Test and Evaluation Policy</i>	http://www.apd.army.mil/jw2/xmldemo/r73_1/head.asp
<i>Electronic Code of Federal Regulations; Title 38: Pensions, Bonuses, and Veterans' Relief; Part 1—General Provisions; Program Evaluation</i>	http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=8b37d430a012029310f51ea9da70644d&rgn=div8&view=text&node=38:1.0.1.1.2.0.4.4&idno=38
<i>Title 38 – Veterans' Benefits; Part I- General Provisions; Chapter 5 Authority and Duties of the Secretary; Subchapter II- Specified Functions; Sec. 527 - Evaluation and data collection</i>	http://www.gpo.gov/fdsys/search/pagedetails.action?browsePath=Title+38%2FPart+I%2FChapter+5%2FSubchapter+II%2FSec.+527&granuleId=USCODE-2010-title38-partI-chap5-subchapII-sec527&packageId=USCODE-2010-title38&collapse=true&fromBrowse=true&bread=true
Process Improvement	
Deputy Chief Management Officer Continuous Performance Improvement (CPI)	http://dcmo.defense.gov/products-and-services/continuous-performance-improvement/
Army Performance Improvement Criteria (APIC)	http://www.hqda.army.mil/leadingchange/APIC/APIC%202002/2002%20APIC%20PDF.pdf
Air Force e-Publishing Process Improvement documents	http://www.e-publishing.af.mil/?rdoFormPub=rdoPub&txtSearchWord=PROCESS+IMPROVEMENT&client=AFPW_EPubs&proxystylesheet=AFPW_EPubs&ie=UTF-8&oe=UTF-8&output=xml_no_dtd&site=AFPW_EPubs&btnG.x=11&btnG.y=11
Army Office of Business Transformation Continuous Process Improvement Knowledge Center	http://www.armyobt.army.mil/cpi-kc-welcome.html
Veterans Health Administration; Systems Improvement Framework (Guideline)	http://www.paloalto.va.gov/docs/ImprovementGuide.pdf
Research with Human Subjects	
AR 70-25. Use of Volunteers as Subjects of Research.	http://www.apd.army.mil/jw2/xmldemo/r70_25/cover.asp
DODD 3216.2. Protection of Human Subjects and Adherence to Ethical Standards in DOD Supported Research.	http://www.dtic.mil/whs/directives/corres/pdf/321602p.pdf
Health and Human Services	http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html

Organization and Title	Source
U.S. Food and Drug Administration Center for Biologics Evaluation and Research (CBER) Guidance for Industry E 10 Choice of Control Group and Related Issues in Clinical Trials	http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073139.pdf
U.S. Food and Drug Administration ICH Guidance Documents	http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/GuidancesInformationSheetsandNotices/ucm219488.htm
Department of the Navy Human Research Protection Program	http://www.med.navy.mil/bumed/humanresearch/resource/Pages/DO_NHRPPGuidance.aspx
Navy Regulations, Directives, Instructions and Policies for the Protection of Human Subjects in Research	http://www.usna.edu/HRPP/docs/Enclosure1_REGULATIONSANDPOLICIES_references_SGapproved_Aug07.pdf
DOE Openness: Human Radiation Experiments: Chapter 3: The Development of Requirements for Human Subject Research in Other Federal Agencies	http://www.hss.doe.gov/healthsafety/ohre/roadmap/achre/chap3_3.html
Electronic Code of Federal Regulation; TITLE 38--Pensions, Bonuses, and Veterans' Relief; Chapter I—Department of Veterans Affairs; Part 16—Protection of Human Subjects	http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=8b37d430a012029310f51ea9da70644d&tpl=/ecfrbrowse/Title38/38cfr16_main_02.tpl
VHA Directive 2003- 031; Establishment of a Facility Human Protections Program	http://www.research.va.gov/pride/policy/default.cfm#directives
VHA Directive 2003- 065; Accreditation of Human Research Protection Programs	http://www.research.va.gov/pride/policy/default.cfm#directives
VHA Handbook 1058.03; Assurance of Protection for Human Subjects in Research	http://www.research.va.gov/pride/policy/default.cfm#directives
VHA Handbook 1200-05; Requirements for the Protection of Human Subjects in Research	http://www.research.va.gov/pride/policy/default.cfm#directives
Evaluation and Inspection	
Army Regulation 70-74: Independent Research and Development	http://www.apd.army.mil/jw2/xmldemo/r70_74/head.asp
Army Regulation 1–201: Army Inspection Policy	www.apd.army.mil/pdf/files/r1_201.pdf

APPENDIX IV: GUIDE TO CONDUCTING INTERVIEWS AND FOCUS GROUPS FOR PROGRAM EVALUATION

An essential program evaluation function is to collect and analyze data pertaining to the effectiveness and implementation of a program's objectives. Interviews and focus groups are reliable qualitative data collection tools that are among the most widely used for program evaluation. They are often employed when detailed information about individuals' perspectives are necessary. Though these methods generally require more time and resources than other widely used tools, they provide valuable insight and context into a program that can contribute meaningful awareness to quantitative data. Interviews and focus groups are similar methods, yet distinct in that an interview gathers data on the perspectives of one person, while a focus group gathers information from a group of people. This brief section will offer an introduction to both of these methods and how to employ them effectively in the context of program evaluation.

Interviews

Interviews are a useful tool that can yield detailed information on stakeholder opinions regarding a program's efficiency, impact or general feelings of satisfaction. These are typically one-on-one meetings between an evaluator and a participant that are useful in acquiring stories and experiences which give context to quantitative data. Interviews may be preferred over focus groups when isolating individual perspectives regarding a program is desired or when it is believed that participants may not feel comfortable sharing their thoughts in a group setting.

Note: Prior to beginning with the interview process, it is important to ensure that Institutional Review Board (IRB) requirements concerning human subjects are being met. While program evaluation protocols are typically exempt from review, it is important to ensure that all procedures all followed in compliance with IRB regulations.

Planning an Interview

Planning interviews requires extensive consideration of the many factors that influence the interview outcome. As with any evaluation tool, it is necessary to assess and control for variables that may affect the validity and reliability of the findings. Proper interview design and planning will help ensure consistency in results. Below is a brief description of the process for developing and conducting effective interviews.

Sources of Information

One of the first phases of planning the interview process is to select participants who represent a diverse range of stakeholders and their views. These stakeholders may include decision makers, staff, patients, health professionals, and so on. Thorough interviews generally seek to acquire a varied pool of sources; however, this factor may be dependent on the type of evaluation being conducted and the types of questions being addressed. A common rule of thumb to consider when determining a sample size for interviews is that enough feedback has been gathered when analogous information—such as issues or themes—is being repeated from various sources. Nonetheless, always ensure that precise data collection measures conform to the parameters of the study.

Interviewer and Setting

One of the strengths of interviews as a data collection method is the ability to acquire information from people in a space where they feel most comfortable sharing their perspectives. Accordingly, it is important to ensure that both the environment chosen and the interviewer's experience help facilitate these results. The setting chosen should be one that is comfortable to the interviewee and conducive to candid discussion with the interviewer, who should be skilled in employing interviewing techniques. The following are some key points to bear in mind when shaping the context of the interview:

- Interviews should be held in a nonthreatening, private setting
- Facilities should be accessible to persons with disabilities
- Interviewer should ensure their body language communicates interest
- Interviewer should avoid language that conveys bias, such as leading statements, or personal opinions

Confidentiality

Confidentiality is one of the key facets of interviewing. Participants should be assured that the input they share will be kept confidential. If choosing to quote a participant, ensure that they have provided consent and that necessary precautions have been taken to prevent their comments from becoming affiliated with their identity. One common measure of confidentiality is to assign each participant a code that replaces their name as an identifier.

Interview Format

The data collection process may be examined in three phases: the introduction, the question segment, and the closing. Each phase of this process should be properly managed to ensure reliability and consistency. A prescribed protocol in the form of an interview guide serves to steer the interview and ensure uniformity throughout implementation.

An interview guide offers the interviewer a set direction on how to communicate and develop rapport with the participant through various stages of the interview process. Beyond the three phases we have discussed, this may also include instructions on key functional roles, such as whether to take notes during an interview or record observations after each interview. This consistency helps increase the reliability of results, facilitates later data analysis, and reduces the probability of interviewer bias.

Introduction

The introduction of an interview is an important segment that informs the interviewee of the study's objective and their role in the process. Below are recommended guidelines for conducting the initial phase of an interview.

- Explain to the interviewee the purpose of the research, why they were selected, and the anticipated time frame for the interview
- Assure interviewee that confidentiality will be maintained and specify any processes that have arranged to ensure this (e.g. assigning codes)
- If recording the interview, clarify the purpose of the tape recording and request their permission to proceed using the tool. If a note taker is present, identify the person and explain their function
- Seek the participant's consent, whether verbal or written
- Notify the interviewee that they are not obliged to respond to questions and may stop the interview at any point

Questions

Questions should be framed according to the type of evaluation being conducted (process, outcome, or impact) and the larger question the evaluation is trying to answer. Structuring these questions through an interview guide is important to helping maintain consistency across interviews. If interviews are targeted at various stakeholder groups, multiple guides may be necessary since questions will likely differ for each group. Below are useful strategies to help with the question development process.

Tips for Developing an Interview Guide:

- *Ask open-ended questions as opposed to close-ended questions.* For example, in place of asking, “Are you familiar with the services this program offers?” ask “What services offered by this program are you familiar with?”
- *Ask factual questions before opinion questions.* For example, consider asking, “What types of services did you receive through this program?” prior to asking, “What did you think of the services you received?”
- *Use probes to keep the conversation flow and acquire individual insight.* Examples include: “Was there anything you wanted to add?” “Could you provide an example?” “How did you respond to that?” “What did you think of it?” These types of questions help steer the conversation and prompt the informant’s perspective without introducing bias.
- *Allocate a time frame to each question to ensure that each question is addressed within the designated scope of time.*

Closing

As part of the interview protocol, the interviewer should have a prescribed outline of what to say when concluding the interview. This entails thanking the interviewee for their participation, and answering any questions they have regarding the process. Allow for final comments they may have, and then inform them of the next steps in the process. This may include offering to provide a synopsis of findings once the evaluation has been concluded.

Managing the Data

Interviews generate an extensive amount of valuable qualitative data, which can be difficult to effectively analyze if not managed using appropriate methods. Following the conclusion of each interview, it is recommended that notes be recorded as soon as possible. Try searching for themes or patterns that can be observed broadly across interviews. Arrange key information based on these patterns by cutting and pasting electronically or using index cards. This method, known as *categorization*, allows broader themes to be assessed against key concepts. **Appendix VIII: Guide to Coding Qualitative Data** contains information on how to code qualitative information from interviews. When using this or any other method of data analysis, the data must always be held in context of the program so that key information is not missed or misinterpreted.

Focus Groups

Focus groups are another useful qualitative data collection tool that provide in-depth information and insight into stakeholder perspectives. Much like interviews, focus groups are useful for surveying the range of opinions and feelings surrounding a topic or a program. Each tool has strengths and weakness that should be measured against the needs of the evaluation and the resources available. Focus groups are typically longer than individual interviews, however take less time and are more economical overall because they gather input from multiple individuals at once. They are a good option when confidentiality and depth of perspective are not of concern. The techniques for conducting focus groups and interviews

are similar. Basic instructions on the process of developing and conducting effective focus groups are provided.

Note: As with interviews, IRB regulations regarding research participants must be considered and fulfilled prior to conducting focus groups.

Planning a Focus Group

As with other data collection instruments, focus groups must be properly designed to ensure that variables which might affect validity or reliability are well-managed. Focus groups require thorough preparation and logistical planning to coordinate multiple participants' schedules, find a skilled facilitator, and avoid common pitfalls.

Sources of Information

Selecting participants for a focus group requires an assessment of the kind of questions and information being sought. Depending on the purpose of the evaluation, it might be beneficial to include various target groups or concentrate on one segment of stakeholders. If evaluating a broad target population, it is necessary to ensure that different stakeholder groups are separated by session, as mixing groups may restrain participants from responding candidly. The size of a focus group can range from six to 12 people, a size large enough to produce a breadth of responses while still allow individual participation. Evaluators should seek to conduct enough sessions to eventually generate a pattern among responses. Numerous sessions will contribute to the reliability of the evaluation—particularly when dealing with more complex issues. However, too many sessions can be redundant and inefficient, producing little distinctive data, while consuming time and resources.

Choosing a Facilitator

Focus groups are often challenging to manage because participants can easily diverge from the objectives and control the course of the discussion. Furthermore, without the necessary moderating skills, the facilitator can inadvertently introduce biased language or opinions that may influence the conversation. Both issues can weaken the credibility and reliability of information gathered. Accordingly, it is essential—more so than with interviews—to designate this responsibility to a reliable and experienced facilitator. An ideal facilitator should have a general understanding of the stakeholder population and sensitivity to issues which are important to participants. It may be practical to use an internal staff person as facilitator given their inherent familiarity with a program; yet, this personal affiliation may leave much room for bias. When possible, a professional facilitator is preferred due to their experience and implicit objectivity. Regardless of whether an internal or external facilitator is used, the general considerations which are included below should be taken into account.

Facilitator tips for conducting a focus group

- *Be familiar with key ideas and objectives.* A thorough understanding of the purpose and expectations of the session will help ensure that key issues are covered throughout the discussion. Familiarity with the focus group guide is also important for ensuring efficiency and focus throughout the focus group.
- *Establish ground rules.* Ground rules—such as allowing others to speak without interruption, being specific and brief when providing examples and giving priority to those who have not spoken—are principles that help shape the course of the discussion and keep participants accountable to basic expectations.
- *Manage time effectively.* Responses can easily extend beyond the scope of what is necessary, particularly where experiences and anecdotes are involved. The facilitator

is responsible for guiding conversations effectively, transitioning to a different question or using probes to acquire further detail when necessary.

- *Encourage broad participation and inclusion.* It is common in such group settings for a few individuals to dominate or steer the course of a discussion. Use probes and other verbal prompts to elicit responses from others participants—to the extent that they are comfortable—and to ensure that the conversation stays focused on the objectives.
- *Listen attentively and show active engagement.* Listening, and showing that you are listening, are equally important responsibilities for a facilitator. Listening attentively during the session helps to recognize whether objectives are being met and effectively employ probes as needed. Showing that you are listening, through gestures like maintaining eye contact or nodding in affirmation, establishes rapport and makes participants more likely to engage.

Focus Group Setting

As with interviews, focus groups should be held in an environment where the participants and facilitator feel comfortable engaging in an open discussion. A quiet, private setting should be arranged that is easily accessible to participants. Arrange the time and location with sufficient advance notice, taking into account the times of the day that are most accommodating for participants.

Confidentiality

Focus groups, by their nature, lack a degree of confidentiality because participants are brought together in an environment where personal opinions and ideas are openly shared. Nonetheless, preserving participant confidentiality throughout the research process is an imperative facet of evaluations. Evaluators should ensure that personal information and remarks made during the session are protected and never publicly released. As specified in the above “Interview” section, a common method of maintaining participant confidentiality is to assign each participant a code that replaces their name. These codes are matched to participant names on a separate sheet that should be kept in a secure location.

Prior to commencing a session, participants should be briefed on expectations of confidentiality. If confidentiality cannot be kept for some reason—as is sometimes the case with sensitive issues—then the facilitator should explain this prior to proceeding.

Focus Group Format

The progression of a focus group can be segmented into three phases: the introduction, the question segment, and the closing. Developing a protocol to facilitate these phases, such as a focus group guide, will contribute to greater consistency and reliability.

Introduction

The introduction of a focus group session is a critical stage that outlines the objectives of the focus group and helps shape participants’ expectations from the outset. This will not only instill a sense of direction, but helps sets the tone for the focus group. Below are recommendations for an effective introduction.

- Inform participants of the purpose of the study, why they were selected, and the goals for the session

- Clearly communicate expectations of confidentiality and any exclusions to this that may exist; a verbal or written consent to proceed should also be acquired
- If recording the session, notify participants of the recording, clarify how it will be used, and request their permission to proceed. If a note taker is present, identify the person and explain their role
- Establish ground rules for conducting the session (see “Facilitator” section for details)
- Inform participants that they are not obliged to respond to questions and may refuse to proceed at any point

Questions

Although the questions should be established much earlier in the program evaluation process, shaping them to fit the context of a focus group setting is important for avoiding many of the pitfalls common to focus groups. Notably, it is useful to develop a focus group guide that will facilitate the course of the discussion. This focus group interview protocol helps the facilitator manage the group dialogue and establishes a consistent foundation for upholding the reliability of the evaluation. The focus group guide should consist of the interview questions—and suggested probes or follow-up questions that help prompt further detail or clarification from participants.

Tips for developing a focus group guide:

- **Ask a simple opening question to help facilitate discussion.** As the focus group progresses, questions should become narrower, particularly when addressing key questions.
- **Include a pre-set list of probes and follow-up questions in the focus group guide.** Thoughtfully developed probes can help prompt further detail or clarification from participants when needed.
- **Be flexible with the sequence of questions.** The facilitator should be well-informed of key questions and objectives for the focus group. If participants discuss a key point early on, the facilitator should allow the order of the questions to proceed logically rather than forcing the discussion in the pre-set direction.
- **Review questions for common phrasing errors.** Poorly structured questions can limit or influence a participant’s answer, increasing the likelihood of bias. Types of questions to avoid include:
 - **Leading questions.** These question structures can prompt or trigger the participant’s opinion or views regarding a certain topic.
 - **Questions that establish a group hierarchy.** Asking participants about their experience or expertise in a given area can inadvertently establish a pecking order which influences group dynamics.
 - **Why questions.** Questions that begin with “why” convey an inflammatory or interrogating tone, which may cause participants to respond defensively.
 - **Double barreled questions.** Avoid questions that structurally appear to address a single issue, yet actually ask two questions at once.
- **Ensure that each question has a designated timeframe.** If the session is running behind, proceed to cover the most important topics in the remaining time.

Closing

The final segment of the focus group should follow as a logical conclusion to the preceding discussion. The facilitator should summarize perspectives and opinions shared regarding key issues, while allowing participants to affirm or comment on the synopsis. Inform participants of

the next steps in the process and open the floor to any final questions. If possible, offer to provide a synopsis of findings once the evaluation has been concluded. Most importantly, thank participants for their time and contribution.

Managing the Data

After collecting sufficient data from which to draw conclusions, it is important to manage and organize that data in a meaningful way. Planning a systematic analytic strategy prior to beginning the evaluation can help ensure consistency and reliability in the findings. Encouraging the facilitator to record their notes immediately after each session helps capture critical observations and thoughts while still fresh in their mind and helps avoid later confusion between sessions. Once focus groups have concluded, a systematic analysis of the data gathered should follow. **Appendix VIII: Guide to Coding Qualitative Data** contains information on how to code qualitative information from focus groups.

Suggested Reading and References

Interviews:

Gelmon, S. B., Foucek, A., & Waterbury, A. (2005). *Program evaluation: Principles and practices* (2nd ed.). Portland: Northwest Health Foundation.

Boyce, C. & Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input*. Pathfinder International.

W.K. Kellogg Foundation. (1998). *Evaluation handbook*.

Focus Groups:

Gelmon, S. B., Foucek, A., & Waterbury, A. (2005). *Program evaluation: Principles and practices* (2nd ed.). Portland: Northwest Health Foundation.

OMNI. (n.d.). *Toolkit for conducting focus groups*.

Simply Better! (n.d.). *Customers in focus: A guide to conducting and planning focus groups*.

APPENDIX V: SOURCES OF EFFECTIVENESS MEASURES

Organization or Tool	Source
Child Healthcare Quality Toolbox	http://www.ahrq.gov/chtoolbox/measure9.htm
The Center for Outcome Measurement in Brain Injury	http://www.tbims.org/combi/list.html
Measuring Outcome in Traumatic Brain Injury Treatment Trials: Recommendations from the Traumatic Brain Injury Clinical Trials Network	http://www.tbi-ct.org/pdf/Bagiella_JHTR_%20March%202010.pdf
The International Brain Injury Association	http://internationalbrain.org/
Defense and Veterans Head Injury Center (DVBIC) Concussion/TBI Screening Page	http://www.pdhealth.mil/TBI.asp#fam
National Quality Measures Clearinghouse	http://www.qualitymeasures.ahrq.gov/browse/by-topic.aspx
Veteran's Health Administration Mental Health Program Evaluation Technical Manual	http://www.rand.org/pubs/working_papers/WR682.html
Joint Commission on Accreditation in Health Care Hospital Based Inpatient Psychiatric Services Core Measurement Set	http://www.jointcommission.org/assets/1/6/HBIPS.pdf
Mental Measurements Yearbook	http://www.unl.edu/buros/
Center for Quality Assessment and Improvement in Mental Health	http://www.cgaimh.org/NIMHQM.htm
National Quality Forum	http://www.qualityforum.org/Measures_List.aspx
Medical Outcomes Trust	http://www.outcomes-trust.org/instruments.htm
Outcome Measurement in VHA Mental Health Services	http://www.va.gov/VATAP/docs/OutcomeMeasMentalHealthStatus2002tm.pdf
American Psychiatric Association Handbook of Psychiatric Measures	http://www.appi.org/SearchCenter/Pages/default.aspx?k=handbook%20of%20psychiatric%20measures
National Institute of Neurological Disorders and Stroke (NINDS), part of the National Institutes of Health	http://www.commondataelements.ninds.nih.gov/TBI.aspx#tab=Data_Standards

APPENDIX VI: INFORMATION ON RECOMMENDED MEASURES FOR PH AND TBI

Measure	Domain	Brief Overview	Administration information	Scoring methodology	Score interpretation	Recommended by
Dizziness Handicap Inventory (DHI)	Balance	DHI is a 25-item questionnaire for patients with vestibular problems that assesses the level of self-perceived handicap. It addresses 3 domains: functional, emotional and physical.	<p>Who Administers? Anyone. Self-administered</p> <p>Who can score? Anyone. However, clinicians should review</p> <p>How long does it take? ~5 to 10 minutes</p>	Each of the three potential responses in the questionnaire has an ascribed score, with the first selection receiving a 0, the second receiving a 2 and the third receiving a 4. The final score is the sum of all selections.	Potential scores lay between 0-100. 0-30=mild; 31-60=moderate; 61-100=severe handicap	OASD memo*
Epworth Sleepiness Scale (ESS)	Sleep	ESS is an 8-item self-report questionnaire that assesses a person's level of daytime sleepiness by examining their probability of falling asleep doing common daily activities.	<p>Who Administers? Anyone. Self-administered</p> <p>Who can score? Anyone. However, clinicians should review</p> <p>How long does it take? ~2 to 3 minutes</p>	Respondents rate their propensity to fall asleep in the context of certain daily situations on a 4-point scale. The sum of all 8 scores is the final score.	Scores may range from between 0 and 24; a higher score indicates a greater level of daytime sleepiness	OASD memo*
PCL-M Questionnaire	PTSD	The PCL is a 17-item self-report questionnaire that directly assesses the DSM-IV symptoms of PTSD. This tool assists with screening, diagnosing or monitoring PTSD symptom changes throughout treatment, among other purposes.	<p>Who Administers? Anyone. Self-administered</p> <p>Who can score? Anyone. However, clinicians should review</p> <p>How long does it take? ~5 to 10 minutes</p>	Each question is scored 1 through 5. Total score is the sum of the 17 items.	The range of scores lay between 17-85. Generally, scores higher than 40 indicate a need for further evaluation.	OASD memo* and both the PTSD and Post-Deployment Health CPGs

Measure	Domain	Brief Overview	Administration information	Scoring methodology	Score interpretation	Recommended by
The Patient Health Questionnaire (PHQ-9)	Depression	The PHQ-9 is a 9-item self-report questionnaire that serves as a diagnostic tool to identify depression. The nine questions directly correlate with DSM-IV criteria for major depressive disorder (MDD). It may also be used to assess symptom severity and track symptom improvement.	<p>Who Administers? Anyone. Self-administered</p> <p>Who can score? Anyone. However, clinicians should review</p> <p>How long does it take? ~5 to 10 minutes</p>	Each question may receive a score between 0 and 3, with 3 indicating a higher severity of symptoms. The sum of scores for all 9 responses is the total score.	Scores range between 0 and 27. Depending on the method of score analysis, results may be used to identify presence of a depressive disorder or to assess the severity of depression.	OASD memo* and both the MDD and Post-Deployment Health CPGs
Military Acute Concussion Evaluation (MACE)	Cognition	The MACE is an examination tool that allows front-line providers to assess cognitive performance based upon four critical domains: concentration, immediate memory, orientation and memory recall. It also entails a neurological exam and symptom screening.	<p>Who Administers? Any medical provider, from medic/corpsman to physician</p> <p>Who can score? Medical provider</p> <p>How long does it take? ~5 to 10 minutes</p>	The cognitive section of the MACE is scored based on accuracy regarding evaluation for the four domains. The final score is the sum of each section. The neurological section and symptom results receive a qualitative score considered alongside the cognitive score.	The maximum score for the cognitive exam is 30. Scores below 25 are considered clinically relevant and prompt further provider evaluation.	OASD memo*
Neurobehavioral Symptom Inventory (NSI)	Cognitive, affective, somatic, sensory	The NSI is a 22-item self-report questionnaire used to capture patient complaints related to a range of symptoms associated with TBI.	<p>Who Administers? Anyone. Self-administered</p> <p>Who can score? Anyone. However, clinicians should review</p> <p>How long does it take? ~5 to 10 minutes</p>	Each question may receive a score between 0 and 4, with 3 indicating a higher severity of symptoms. The sum of scores for all 22 responses is the total score.	Scores range between 0 and 88, with higher scores indicating greater severity of symptoms.	OASD memo*

* OASD Memo- Office of the Assistant Secretary of Defense, Memorandum for Service Surgeon Generals, Symptom Management in Mild Traumatic Brain Injury, May 30, 2008.

APPENDIX VII: CONDUCTING A FOLLOW-UP EVALUATION

Follow-up evaluations allow a program to assess long term effectiveness, moving beyond simple pre-and post-assessments of program effectiveness. A follow-up assessment is typically defined as contacting program participants who are no longer actively involved with the program to conduct an assessment. This type of assessment can provide valuable feedback on the activities or service provided by the program.

Feedback from follow-up evaluations can be leveraged to improve the program as it informs program managers whether the interventions last, allowing them to make changes if effects are not sustained to a great enough degree. Another benefit to conducting a follow-up evaluation is that it maintains a relationship with past participants. (e.g. Follow-up discussions can be seen as booster sessions to reinforce interventions or check in on a past patient.)

Conducting a follow-up evaluation should be prioritized and planned early on, due to the logistical and administrative difficulties of conducting this type of assessment. Conducting follow-up evaluations can be time intensive, and a program may have to hire staff or set aside time in existing staff's workload to maintain this task.

In conducting a follow-up, there are six key questions to consider:

1) *What will be done with the information?*

As with program evaluation, it is important to recognize how the findings will be acted upon. Will the data be used to determine which interventions were superior in terms of lasting effectiveness? Will it be aimed at providing rationale for the program due to proven effect? If so, how will findings be presented? Deciding what will be done ultimately with the information will determine the approach for the follow-up evaluation.

2) *What type of information will be gathered?*

Similar to **Step 2: Develop Evaluation Questions**, the purpose of the follow-up will help guide the questions. Depending on whether the goal is to improve service, advance treatment or interventions, or obtain program evaluation data, questions will need to be constructed accordingly. Normally, the information collected at follow-up will closely align to that collected during the time the participant was in the program, allowing for direct comparison on variables like symptom levels, quality of life, etc.

3) *How will the information be analyzed?*

Adding additional assessments will not only impact data collection, but will also likely affect what statistical procedures are used. If the program evaluation is going to include one or more follow-up assessments for participants, this information will need to be accounted for in both the data sampling plan and data analysis plan.

Benefits of a Follow-Up Evaluation

- Provides feedback on long-term effectiveness of interventions
- Provides rationale regarding impact and need for program
- Maintains a relationship with past participants
- Assists in making the case for the program (to staff, funders, clients, etc.)
- Helps to identify best-practices; reduces need for a costly study because the treatment has already been provided
- Provides data that can be used for future program evaluations

4) How will the information be collected (data gathering) with participants who have left the program?

Because a follow-up occurs once the participants are no longer actively engaged with the program, it is likely that this will need to be done either by telephone or mail. However, if possible, face-to-face encounters allow for personal interaction, flexibility in dialogue, and non-verbal feedback. Needless to say, collecting information from participants who are no longer actively involved with the program may prove difficult. In deciding what method to use, there are several factors to consider:

- Availability
- Cost
- Time to administer and receive results
- Specialized training needs
- Reliability and validity
- Requirements/guidance from higher headquarters or governing agencies

Error! Reference source not found. describes the pros and cons of different options for gathering follow-up data.

Figure 34. Pros and Cons for Data Gathering Methods

Method	Pros	Cons
Face-to-Face	<ul style="list-style-type: none"> • Shows personal interest • Allows for flexibility in conversation • Provides non-verbal feedback • Allows for an impromptu assessment of the patient while also gathering follow-up information 	<ul style="list-style-type: none"> • Scheduling times both parties can meet • Travel may be an issue if the participant does not live near the facility • Cost • Time
Telephone	<ul style="list-style-type: none"> • Does not involve travel • Allows for flexibility in conversation 	<ul style="list-style-type: none"> • Limits contacts to those with phones • Does not allow for nonverbal communication
Mail	<ul style="list-style-type: none"> • Eliminates scheduling issue • Requires the least commitment from the client • Least expensive 	<ul style="list-style-type: none"> • Client may not feel pressure to respond • Questionnaires can be lost or forgotten • Does not allow for nonverbal communication • Does not allow for flexibility in conversation • Reading and writing requirements

5) When will follow-up contacts be conducted?

Typically, follow-ups are conducted at either three, six, and/or 12 months post program completion. The population the program serves and the resources available may help determine when would be best suitable to the program’s needs. Furthermore, the purpose of the follow-up evaluation will may help guide this decision.

6) Who will conduct the follow-up and contact participants?

In considering who will conduct the follow-up, there are a few things to keep in mind. First, it depends on whether the contact will be personal (telephone or face-to-face). If so, one

possibility is having a staff member with whom the participant has interacted and developed a rapport conduct the follow-up. This presents both advantages and disadvantages. It may be beneficial to have someone with a background similar to and an understanding of the participant conduct the interview. On the other hand, it may pressure the participant to provide responses that they think are pleasing to the interviewer. To minimize such bias, having a staff member who has not had contact with the participant is an option. Although this may prevent bias, unfortunately, it may not facilitate participant openness. Another option is to obtain information via a mailed questionnaire, which would eliminate distortion caused by personal interaction. The following table describes the pros and cons of various strategies for contacting past participants. **Figure 35. Pros and Cons for Contact Methods** lists the different options with pros and cons for each.

Figure 35. Pros and Cons for Contact Methods

Contact	Pros	Cons
Staff who interacted with the client while at the program	<ul style="list-style-type: none"> Establishes rapport Familiarity with the case First hand feedback for staff on the care they provided and its impact 	<ul style="list-style-type: none"> Pressure to please the interviewer Staff may have personal bias and interpret the client's report differently
Staff who never interacted with the client	<ul style="list-style-type: none"> Will not sway the client to provide feedback in a positive light Staff member will be less likely to introduce bias or interpret the client's report differently 	<ul style="list-style-type: none"> Client may not feel at ease with the staff member they have never met
Self-administered questionnaire (no interaction with staff)	<ul style="list-style-type: none"> Eliminates distorted results due to client-staff interactions Less expensive Multiple choice format avoids the issue of deciphering handwriting 	<ul style="list-style-type: none"> Less flexibility for expanding on issues or clarifying responses Reading and writing abilities of client

Difficulties of Conducting Follow-Up Assessments

Potential difficulties in contacting participants who are no longer actively involved in the program include scheduling, traveling and lack of interest on the participant's part. Adding in these types of assessments may make the program evaluation more likely to be deemed research by an IRB. It is important to consult with an IRB representative to inquire about whether the plan for follow-up will require IRB approval. Even if the program evaluation is not deemed to be research, conducting a follow-up requires special care in terms of protecting confidentiality. Potential program participants slated for follow-up should have signed an informed consent for later contact. This informed consent should detail the rationale for the evaluation, what types of questions will be asked, and policies for maintaining confidentiality.

Suggested Reading and References

Wolfe, B.L. & Miller, W. R. (n.d.). *Program evaluation: A do-it-yourself manual for substance abuse programs*. Albuquerque, NM: University of New Mexico, Department of Psychology.

APPENDIX VIII: GUIDE TO CODING QUALITATIVE DATA

Introduction to Coding Qualitative Data

As discussed in **Step 6:**, qualitative data are descriptive data (words, phrases, sentences and paragraphs), most often collected through techniques such as interviews, observations, journaling, etc. It is difficult to quantify or simplify these data to a numerical value as it includes abstract concepts such as emotions, opinions, or experiences. As such, qualitative data requires a distinct data analysis process than that which is used for quantitative data. To facilitate analysis, qualitative data is often *coded* by assigning portions of data to established categories.

The *coding* (or *categorization*) process can be described in five-steps. A description of each step is provided in **Figure 36. Steps to Coding Qualitative Data for a Program Evaluation**, and described more fully below.

Figure 36. Steps to Coding Qualitative Data for a Program Evaluation

1. Become familiar with the data that will be coded. This may be through reading transcripts or reports, listening to audio-taped interviews or interactions, etc. The goal of this step is to learn as much as possible about the data to facilitate the coding process.
2. Identify categories for coding. These categories may be informed by the known goals or objectives of the program, the purpose of the evaluation, themes identified in the data, etc. Categories identified <i>before</i> coding begins are called preset categories while categories that are developed as a result of the coding process are generally called emergent categories.
3. Code the data. Assign values to the categories and review the qualitative data for evidence of these categories or themes. Remember that coding is not a static process and that categories may need to be added or collapsed depending on the level of detail and themes that are identified in a data set. In the early stages of the coding process, it is important to create and maintain a detailed codebook.
4. Analyze the coded data. Using descriptive and/or inferential techniques, summarize, analyze and discuss the patterns, themes, and differences that are evident in the data set.
5. Interpret the analysis outcomes. Based on the structure and objectives of the evaluation, and the limitations of the data set, determine what conclusions can be drawn from the qualitative data.

- 1. Become familiar with the data that will be coded.** The first step in coding qualitative data is for the coders to familiarize themselves with the dataset. This may be through listening to audio recordings, watching videos, reading transcripts, etc. Familiarity facilitates the major goal of coding which is to condense a plethora of information into specific categories.
- 2. Identify categories for coding.** Having a clear theme or topic for the evaluation questions will help generate a list of categories. Qualitative responses can then be grouped according to this list. These categories or topic areas might be derived from the goals of the program or from a relevant literature search and are sometimes referred to as *preset categories*. Once the preset categories have been identified and defined, a coder or analyst would read through the qualitative responses, searching for information that fits the categories.

Because qualitative data generally result from open-ended or free response questions, the preset categories may not capture the full range of responses and may need to be

augmented. As additional themes are identified, it may be necessary to develop additional categories. These categories are called *emergent categories* and expand the categories into which a coder may classify qualitative data. It is particularly important to keep notes with explanations of emergent categories, as there may be no documentation of them.

The coding process can also be refined in the opposite direction, reducing the number of categories. If the data are more narrow in scope than anticipated, similar categories with low response rates may need to be collapsed. This can be done during the coding process or after coding has been completed. Developing super categories follows a similar process, but with a different objective. *Super categories* are larger categories that reduce the number of categories by collapsing across common themes. They can help evaluators identify patterns and connections that do not fall specifically within one category, and are therefore easily missed if coders do not consider the dataset as a whole.

- 3. Code the data.** As it is informed by the evaluation's goals and objectives as well as the data, coding is not a static process. Therefore, it is imperative that coders keep careful documentation of the procedures and definitions used, and decisions made while coding. As coding transforms the qualitative data into a more quantitative or empirical form for analysis, handling of data and process replicability should meet the high standards and rigor required of statistical analysis.

There are several procedures to use during the coding process to ensure rigor. First, a clear and descriptive codebook should be kept. The codebook may contain information about: (1) categories; (2) their descriptions and definitions; (3) the criteria used to define categories and place data within them. Information needs to be detailed and specific. In theory, someone unfamiliar with the evaluation should be able to use the information contained in the codebook to accurately code qualitative data from the study. The coding key should also be within the codebook. Remember, the purpose of coding qualitative data is to transform it into a format that can be analyzed empirically or statistically or presented descriptively. The key should describe the categories being coded and the values assigned to them.

Second, use multiple coders as a check on the reliability of the coding process. Although multiple coders add cost and time to a study design, they also add greater assurance that the interpretation of the qualitative data required in the coding process is not being influenced by biases or confounders. If budget and deadlines are prohibitive, consider using two coders, but only double coding a selection of the responses to establish inter-rater reliability and verify that the coders are using well-described operational definitions for categories and other coding concepts.

- 4. Analyze the coded data.** Once coding process has been completed and inter-rater reliability checks, the dataset will be ready for analysis through descriptive or inferential statistics. As discussed in **Step 6:**, *descriptive statistics* include a series of techniques that provide a summary or a "big picture" view of the data. Descriptives are a useful first step in data analysis because they can identify overarching themes or trends in the data that can help guide later analyses. *Inferential statistics* provides a means to interpret data and draw conclusions that might be generalizable to larger populations by comparing different groups. For example, when coding for certain behaviors, the number of a certain behavior exhibited by individuals who participated in a program could be compared with the number of that behavior exhibited by individuals who did not.

- 5. Interpret the analysis outcomes.** At the end of the coding process, discussion and conclusions should be able to be drawn to address the evaluation goals, objectives, and questions.

An example case is provided below in **Figure 37. Example of Coding Qualitative Data.**

Figure 37. Example of Coding Qualitative Data

<p>You are conducting an evaluation of a program that provides support for service members reintegrating after deployment. One component of the program matches each service member to a mentor who provides support, advice and referrals to other resources during the reintegration process. As part of your program evaluation, you are interested in finding out how the mentors are perceived by the service members and whether the service they provide is viewed as valuable. Interviews with service members, including questions about their mentors, were audio taped and now need to be coded and scored.</p>
<p>1. Familiarize yourself with the data that you will be coding. As the coder, you request a copy of the interview questions. You also listen to a sampling of the audio-taped interviews to familiarize yourself with the format of the interview and the general length of responses. While reviewing the questions and the tapes, you begin to identify important themes that should be addressed by the coding process.</p>
<p>2. Identify categories for coding. Based on the goals of the program, the evaluation team has identified several categories into which responses may fit. They would like you to code responses to three questions from the interview for positive and negative comments about the mentors. Based on your initial review of the tapes, you noticed that some responses were not clearly positive or negative, so you also add a neutral category.</p>
<p>3. Code the data. You assign values to the categories (positive = +1; negative = -1; neutral = 0) and note these assignments in your code book. You also include information about how to classify a response as positive, negative or neutral and provide several examples of each. You listen to each of the audio-tapes and assign values to statements from participants in response to the three mentor questions. While coding the data, you notice that the majority of comments reference the availability of the mentor. Although this does not fit neatly into one of the identified categories, you note this in your code book as a common theme across respondents.</p>
<p>4. Analyze the coded data. Once all of the interactions have been coded, you perform descriptive analyses to determine the frequency of each category of response as well as a total response score for each participant. You present these data in text and graphical form.</p>
<p>5. Interpret the analysis outcomes. From the analyses, you note that that the majority of responses about the mentors were neutral or negative. You also note that most of the complaints were about a lack of availability. Based on these findings, you conclude that the program should identify ways to make mentors more accessible to the service member and should consider including information about availability in the mentor training course.</p>

Suggested Reading and References

Siedel, John. (1998). *Qualitative data analysis*. Qualis Research.

Renner, Marcus. Taylor-Powell, Ellen. (2003). *Analyzing qualitative data*.

APPENDIX IX: OVERVIEW OF BASIC STATISTICAL ANALYSES

Statistics are tools for translating program data into valuable information for program understanding or improvement. There are various ways to apply statistics, and each method is dependent on the type of question being answered. The clearer the question, the easier it is to define the metrics and analysis that can best answer the question. Questions should be applicable to the type of data the program can collect with the available resources.

Figure 38. Sample Research Questions

<ol style="list-style-type: none"> 1. What psychological health condition is most prevalent within the unit? 2. Is there a change in behavior of service members due to group discussion? 3. Are PTSD symptoms higher in deployed or non-deployed personnel? 4. How many patients does our clinic treat on a weekend compared to a weekday?

Data collected and the execution of an analysis can be geared towards describing a program (*descriptive*) or inferring whether a program action results in an intended or unintended effect (*inferential*). **Figure 39. Descriptive and Inferential Statistics** provides an overview of descriptive statistics and two common types of inferential statistics: associative and comparative.

Figure 39. Descriptive and Inferential Statistics

Statistics Type	Description
Descriptive	Provides preliminary basic descriptive facts for initial exploration of program impact and to develop hypotheses for future analysis.
Inferential	<p>Associative -Determines the relationship among participant data, such as co-occurring conditions</p> <p>Comparative-Reveals statistical difference and significance of participant data</p>

As questions about a program become geared towards providing inference, statistics advance from determining averages and counts to associative and comparative tests. These advanced statistical tests of inference are used to detangle program complexity involving multiple types of data or variables. If necessary, program leaders can work with a statistical professional to assist in conducting complex data analysis for their programs. Even with a good program evaluation study design and assistance of a statistician, selection of the appropriate statistical test and subsequent interpretation of the data can be difficult. **Figure 40. Sample Statistical Tests** lists the tests relevant to each type of statistics.

Figure 40. Sample Statistical Tests

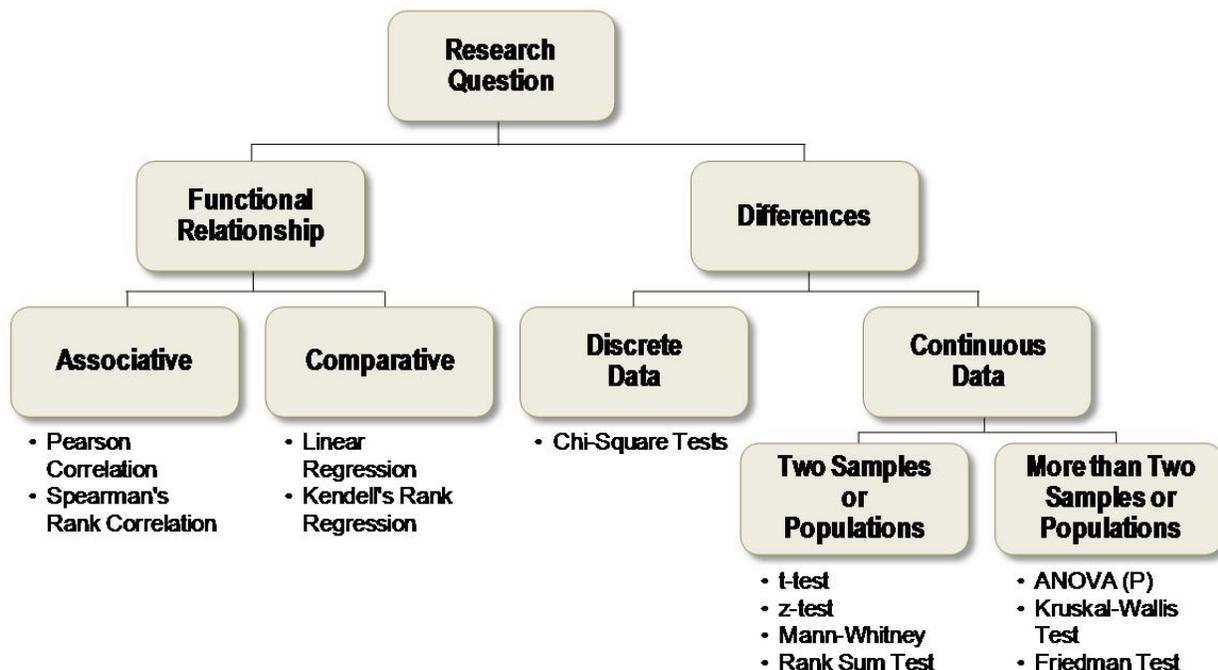
Descriptive	Inferential: Associative	Inferential: Comparative
<ul style="list-style-type: none"> • Mean (average), median and mode • Standard deviation and standard error • Frequency counts • Confidence intervals 	<ul style="list-style-type: none"> • Pearson coefficient • Spearman coefficient • Linear regression 	<ul style="list-style-type: none"> • Paired and unpaired t-test • Mann-Whitney U-test • Analysis of Variance (ANOVA)

How to Select the Appropriate Statistical Test. It is necessary to understand what types of data the program is collecting in order to select the appropriate descriptive or inferential statistical test. Some data are *discrete* or *categorical* values, such as gender, medication type, level of care or satisfaction scores on a numeric scale (i.e., 1= very dissatisfied; 2= dissatisfied and 3= neutral). There is no method when using discrete data to determine whether X is larger, smaller or quantitatively different than Y. X is simply X and not Y.

In contrast, *continuous* variables are values sequenced within a linear distribution, such as height, weight and temperature. Within continuous variables, it can be determined if 9.5 is less than or greater than 10.5. If a variable can be subdivided, it is probably continuous (weight or dollars, for example). Continuous variables can be converted to discrete data (greater than or less than five drinks per week), but discrete variables cannot be treated as continuous. Continuous variables can be classified as nominal, ordinal or interval/ratio.

It is necessary to understand the distinctions between these types of data and the population in which being studied in order to choose between statistical tests. **Figure 411. Statistical Options Flowchart** presents a flowchart of statistical options available for specific data types.

Figure 411. Statistical Options Flowchart

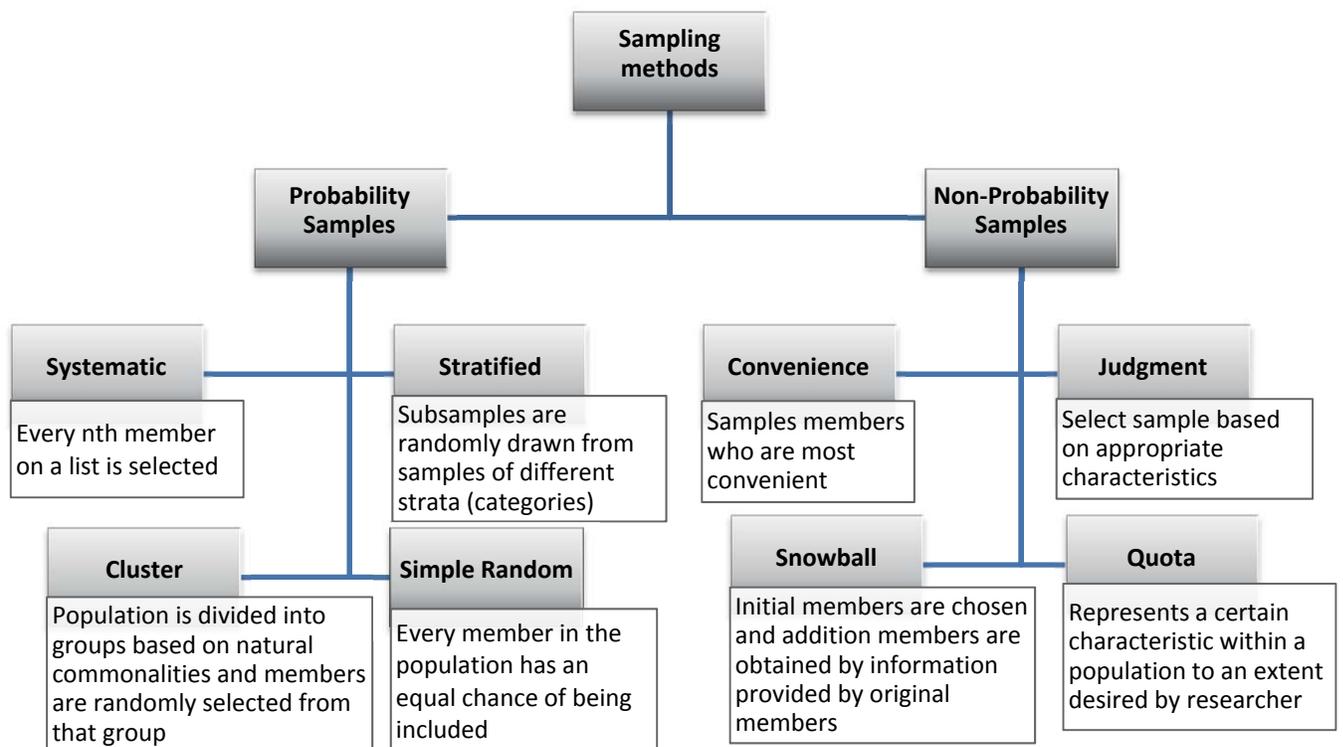


A key factor in statistical analysis of data is ensuring that the data being collected and later analyzed is an appropriate sample of the population for which conclusions will be drawn.

Developing a sampling plan before data is collected can help ensure that the data analysis will accurately reflect the population in question. The first step in sampling is choosing between a probability or non-probability sample. With a probability sample, every person in a population has a known chance of being selected. In a non-probability sample, members are selected based on personal judgment or convenience and not through a statistical technique, which will prevent the results from being generalizable.

Once a decision has been made between a probability or non-probability sample, the next step is to consider the various methods of sampling within each genre. **Figure 422. Sampling Methods** provides a breakdown of the various sampling techniques, with brief descriptions of each.

Figure 422. Sampling Methods



Suggested Reading and References

UCLA Academic Technology Services. (n.d.) Statistical computing. Retrieved February 2, 2012, <http://www.ats.ucla.edu/stat/>.

Fowler, F.J. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

APPENDIX X: TEMPLATES

- Template A: Background Review
- Template B: Mission Statement
- Template C: Goals Template
- Template D: Objectives Development
- Template E: Stakeholder Analysis
- Template F: Stakeholder Engagement Presentation
- Template G: Program Inputs Brainstorming Sheet
- Template H: Program Activities Brainstorming Sheet
- Template I: Program Outputs Brainstorming Sheet
- Template J: Program Outcomes Brainstorming Sheet
- Template K: Logic Model Template
- Template L: Data Sampling Plan
- Template M: Data Monitoring Plan
- Template N: Program Evaluation Report
- Template O: Evaluation Report Feedback Request
- Template P: Evaluation Report Approval Tracking Form
- Template Q: Final Evaluation Report Distribution Letter
- Template R: Results Implementation Worksheet

TEMPLATE A. BACKGROUND REVIEW

This template is intended to guide program leaders through the process of reviewing the program background. This involves conducting a high-level examination of the problem, identifying existing capabilities to address the problem by conducting a leadership discussion of the current program, and comparing it to other similar existing programs. When completing this form, it may be beneficial to use an interview format to gain the best insight into other programs. The questions below provide structure for a program background review.

Questions Used to Review Program Background

Document Desired Result

What is the desired result of this program?

Document Current State

To document the current state, interview program leaders to answer the following questions:

- How and when did the program begin? For example, was the program established as a result of a law, congressional mandate, higher headquarters mandate, or to answer a local area of concern?
- What does leadership perceive is lacking in terms of addressing the identified need or dealing with the problem?
- What are the challenges and successes of this program?
- What similar programs exist which address the area of concern? What lessons learned, if any, can be applied from existing programs?
- What additional services and resources are not currently available to help address the problem?

TEMPLATE B. MISSION STATEMENT

MISSION STATEMENT QUESTIONS

- What problem does your program exist to correct or overcome?
- What need does your program intend to meet?
- What population does your program intend to target?
- What geographic area does your program serve?
- What philosophy underlies your program?
- What beliefs guide the approach you will take to solving the problem you have identified? (i.e., do you focus on cognitive/behavioral skills or medical/rehabilitative treatment?)

Write your mission statement here:

Reference-National Institute on Drug Abuse, NIH Publication No. 95-3609, Printed 1993, Reprinted 1995

TEMPLATE C. GOALS WORKSHEET

This worksheet is provided to assist program evaluators formulate program goals.

- | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none">• What things must you accomplish to achieve your mission?• What changes must you bring about in your client population?• What is the role of the program in terms of the larger community?• What must be done to ensure that the program will have needed resources?• What is the program's role in advancing the state of the science with regard to treatment?• What other things need to be accomplished? |
| <ul style="list-style-type: none">• The goals should you select should identify end results or accomplishments, instead of processes or steps leading to accomplishments.• If the list of goals identified is too long, then divide the list by prioritizing those questions to be addressed now from those questions that may be addressed at a later time. |

Write your final list of priority goals here:

PRIORITY GOALS
Goal 1.
Goal 2.
Goal 3.
Goal 4.
Goal 5.
Goal 6.

Reference-National Institute on Drug Abuse, NIH Publication No. 95-3609, Printed 1993, Reprinted 1995

TEMPLATE D. OBJECTIVES DEVELOPMENT

This template is intended to guide program leaders in formulating SMART objectives. Complete one of these templates for each Goal identified.

Instructions: Use the following table to brainstorm whether the defined objectives meet the SMART criteria. Columns labelled with the SMART criteria are meant to be completed with a Yes/No response. If the response is no, then think about why the objective does not meet the criteria and make the appropriate changes to the objective.

LIST OBJECTIVES HERE	SPECIFIC	MEASURABLE	ACHIEVABLE	RELEVANT	TIME-BOUND
	Detailed Well-Defined	Quantitative Observable	Actionable Attainable	Purposeful Applicable	Timely Begin/End

TEMPLATE E. STAKEHOLDER ANALYSIS

To complete the stakeholder analysis, interview each identified stakeholder group, completing the fields in the form below.

Stakeholder name	Role or relation to program	Use of program materials or services	Areas of concern for this stakeholder	Follow-up meetings planned or requested

Step	Build Support with Stakeholders
Timeline	Low Complexity: Week #1; Medium Complexity: Week#1; High Complexity: Weeks #1-2
Purpose	This presentation is used to make introductions, describe the evaluation methodology, and solicit feedback on stakeholder interest in the evaluation.
Prepared by	Analyst
Recipient	Evaluation participants (namely stakeholders)

[Organization Conducting Evaluation] Program Evaluation Stakeholder Engagement Presentation

[Program Name (Acronym)]

DD MM YYYY

Stakeholder Engagement Presentation

This briefing reviews the [Organization Conducting Evaluation] Program Evaluation of [Program Name].

Agenda

Introductions

[Organization Conducting Evaluation] Background

[Organization Conducting Evaluation] Program Evaluation Overview

[Organization Conducting Evaluation] Program Evaluation Process

Key Dates

Next Steps

Contact Information

Questions

Introductions

[Organization Conducting Evaluation] Staff

[Full Name of Team Lead]	Team Lead for [Program Name] Program Evaluation
[Full Name of team member]	Subject matter expert(s)
[Full Name of team member]	Analyst

[Organization Conducting Evaluation] Background

Mission Statement:

- [Organization Conducting Evaluation] composition
- Organization Goals
- Organization Objectives

[Organization Conducting Evaluation] Program Evaluation Overview

Purpose of the Evaluation:

- Assess the impact of the program on service members, veterans, families, and stakeholders
- Analyze the program's effectiveness measures
- Document the program's accomplishments, lessons learned, and opportunities

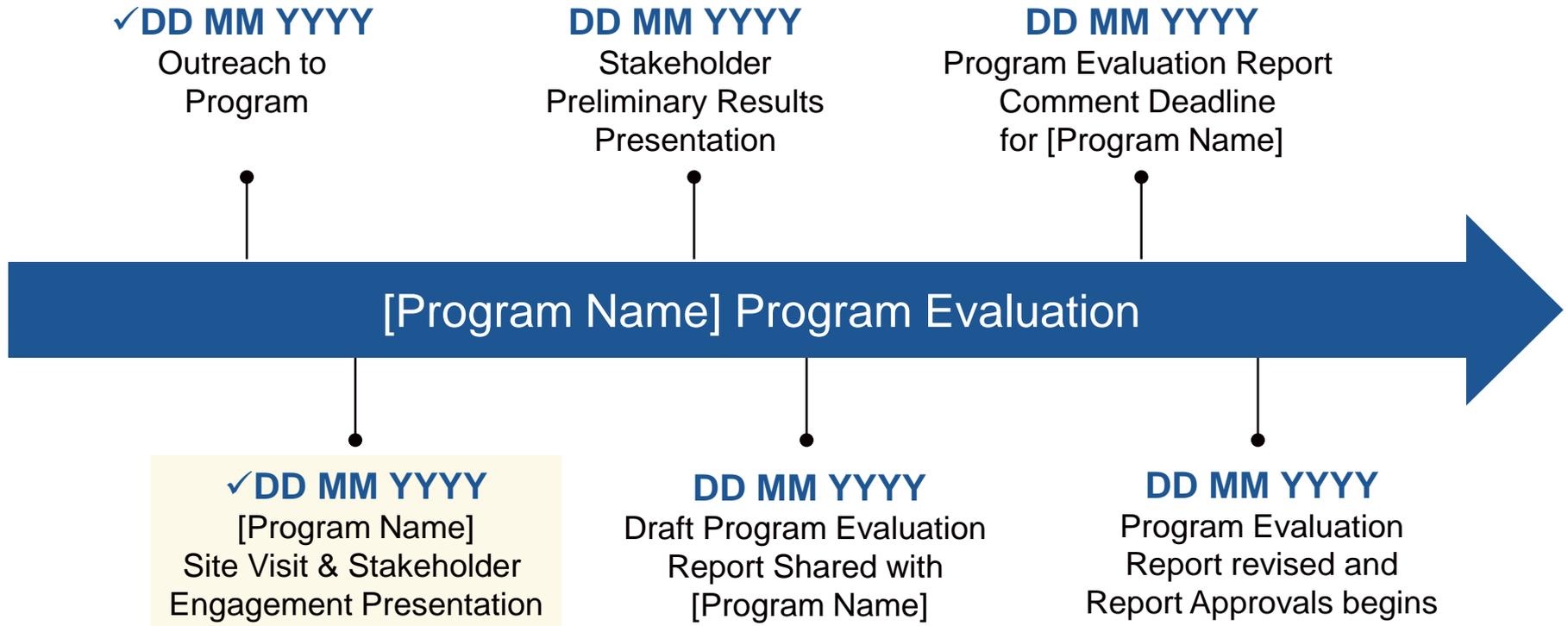
Desired Outcomes and Benefits:

- Apply findings from the Program Evaluation Report to improve the effectiveness of the program
- Support a program and military culture of continuous improvement
- Utilize the Program Evaluation Report to broaden the program's visibility to the intended community

[Organization Conducting Evaluation] Program Evaluation Process

	Preparation	Execution	Feedback
[Organization Conducting Evaluation]	<ul style="list-style-type: none">▪ Provide an overview of the program evaluation process to stakeholders▪ Collect and review available documentation	<ul style="list-style-type: none">▪ Conduct interviews, analyses, and site visits▪ Highlight strengths, lessons learned, opportunities, and performance measures	<ul style="list-style-type: none">▪ Share Program Evaluation Report with stakeholders
[Program Name]	<ul style="list-style-type: none">▪ Share preexisting program data▪ Discuss program structure and processes	<ul style="list-style-type: none">▪ Provide feedback on [Organization Conducting Evaluation's] initial observations	<ul style="list-style-type: none">▪ Collaborate and share insights with intended community▪ Maintain open lines of communication related to major milestones

Key Dates



Next Steps

Stakeholder Preliminary Results Presentation

- [Propose a tentative date for Stakeholder Preliminary Results Presentation]
- Present initial findings of [Program Name]
- Focus on strengths, lessons learned, and opportunities of [Program Name]
- Initial suggestions for future collaborations to increase program effectiveness

Program Evaluation Report*

- **DD MM YYYY:** Draft Program Evaluation Report delivered to [Program Name]
- **DD MM YYYY:** [Program Name] provides Report comments to [Organization Conducting Evaluation]
- **DD MM YYYY:** [Organization Conducting Evaluation] incorporates Reports comments, begins Report approvals process

* Tentative dates

Contact Information

For questions or comments, please contact:

Contact	Role	Email	Telephone
[Full Name of Team Lead]	[Role]	[Email address]	[Phone Number]
[Full Name of team member (if applicable)]	[Role]	[Email address]	[Phone Number]

Questions



TEMPLATE G. PROGRAM INPUTS BRAINSTORMING SHEET

Use this sheet to generate information on the program's resources, which can include staff, space available, levels of funding, and equipment.

STAFF: Full-Time Part-Time Volunteers Type (MD, RN, PT, OT, etc.)	
SPACE: Dedicated Shared	
BUDGET: Salaries Training Travel Marketing	
SUPPLIES / EQUIPMENT: Medical Office	
OTHER:	

TEMPLATE H. PROGRAM OUTPUTS (ACTIVITIES) BRAINSTORMING SHEET

In generating a listing of the outputs for your program, try to consider the all of activities that your staff engage in on a day-to-day basis. The table below lists several example activities, and provides a space for you to document your own programs outputs.

PROGRAM OUTPUTS (ACTIVITIES)		
Treatment Planning Medication Management Diagnostic Testing Patient Follow-Up Calls Field Trips / Outings Patient Education	Individual Treatment Sessions Group Therapy Sessions Workshops Classes Documentation / Charting Interviews	Command Liaison Distributing Materials Developing Public Service Announcements Website Hosting
1)		
2)		
3)		
4)		
5)		
6)		
7)		
8)		
9)		
10)		
11)		
12)		
13)		
14)		
15)		
16)		
17)		

TEMPLATE I. PROGRAM OUTPUTS (PARTICIPANTS) BRAINSTORMING SHEET

In generating the second half of the outputs for your program-(program participants and satisfaction), consider what population you serve, and what kinds of satisfaction measures are in place. Some examples are provided

PROGRAM OUTPUTS (PARTICIPANTS)	
Demographics and other characteristics: <ul style="list-style-type: none"> • Age • Gender • Rank • Service • Component • Typical diagnoses • Common co-morbidities 	Satisfaction: <ul style="list-style-type: none"> • Patient satisfaction surveys (command level) • Patient satisfaction surveys (program level) • Referral sources satisfaction surveys
Demographics & other characteristics:	
Satisfaction:	

TEMPLATE J. PROGRAM OUTCOMES BRAINSTORMING SHEET

In generating a listing of the outcomes for your program, try to consider outcomes that span the immediate (short-term) to distant (long term) time frame. Example outcomes for each are shown below.

PROGRAM OUTCOMES		
SHORT-TERM	MID-TERM	LONG-TERM
Knowledge Levels Skill Levels Attitudes	Decrease in Symptoms Increased Coping Skills Decrease in Risk Factors	Decreased Disability Absence of Disorder Improved Quality of Life
1.	1.	1.
2.	2.	2.
3.	3.	3.
4.	4.	4.
5.	5.	5.

TEMPLATE K. LOGIC MODEL TEMPLATE

<-----Process Evaluations----->

<-----Outcome Evaluations----->

INPUTS	OUTPUTS		OUTCOMES		
RESOURCES	ACTIVITIES	PARTICIPANTS	SHORT-TERM	MID-TERM	LONG-TERM

TEMPLATE L. DATA SAMPLING PLAN

Evaluation Question	What Data Will Be Collected?	How Will Data Be Collected?	From Where Will Data Be Collected?	When Will Data Be Collected?	Who Scores Measure? (If Applicable)
EXAMPLE: Are our patients with an Axis I diagnosis of depression (MDD or Dysthymia) showing clinical improvement?	Beck Depression Inventory- II (BDI-II)	Paper and Pencil Self-Report	<ul style="list-style-type: none"> All patients in program Any patient with depressive disorder 	<ul style="list-style-type: none"> Upon entry to program as part of intake packet At the end of treatment or termination session 	<ul style="list-style-type: none"> Technicians trained in scoring BDI-II must be reviewed by provider (due to suicide question)
1.					
2.					
3.					
4.					
5.					
6.					
7.					
8.					

TEMPLATE M. DATA MONITORING PLAN

This document serves as a template to help guide evaluators through the program evaluation process. The data monitoring template is supplemental to the data collection plan and the sampling strategy template. This document is intended to support establishment of a quality data monitoring plan, and to assist with the collection, documentation, of quality data. This template provides a framework to assist data collection staff and evaluators to ensure the collection and maintenance of quality, valid, reliable data.

DATA MONITORING PLAN	
1. Types of Data	The review of this data on a consistent basis will allow for the overall understanding of the quality of data monitoring and verification checks.
2. Roles and Responsibilities	Evaluators should designate the specific roles and responsibilities of key data collection staff. Consider the following questions: <ul style="list-style-type: none">• Who will verify data accuracy, by what method and how frequently?• Who will verify compliance with the program plan?• How will compliance be verified and how often?• Will a data monitoring committee be formed? Describe the committee (if applicable, include name, credentials, title, organization and contact information of each member).• What are the mechanisms for maintaining independence of judgment?
3. Monitoring Schedule	Develop a schedule that identifies when data will be reviewed, assessed, and verified (i.e. monthly or quarterly).
4. Criteria	Identify the criteria that should be examined to assess data accuracy and validity.
5. Action Steps	Specify the sequential steps that staff should follow to correct any data inconsistencies.
6. Reporting	Develop processes and instructions by which data management staff will perform quality checks in preparation for quality reviews. Detail the process for documenting and reporting and tracking any data inconsistencies. <ul style="list-style-type: none">• What are the timeframes for reporting?• What mechanism will be used to report (specify forms and procedures)?• Who will prepare and submit the report?• How will outcomes be communicated to stakeholders?



DEFENSE CENTERS OF EXCELLENCE
For Psychological Health & Traumatic Brain Injury

TEMPLATE N. PROGRAM EVALUATION REPORT

[Program Name (Acronym)]

Program Evaluation Report

DD MM YYYY

2345 Crystal Drive ★ Crystal Park 4, Suite 120 ★ Arlington, Virginia 22202 ★ 877-291-3263

1335 East West Highway ★ 9th Floor, Suite 640 ★ Silver Spring, Maryland 20910 ★ 301-295-3257

www.dcoe.health.mil ★ Outreach Center: 866-966-1020

Table of Contents

1.0	Executive Summary.....	1
2.0	Program Evaluation Approach.....	1
3.0	Program Summary	1
3.1	Program Need	2
3.2	Program Structure	2
3.3	Program Process	2
3.4	Program Outcome	2
3.5	Program Scalability.....	2
4.0	Program Evaluation Results	2
4.1	Strengths	2
4.2	Lessons Learned	2
4.3	Opportunities	3
4.4	Effectiveness Measures.....	3
5.0	Summary.....	4
6.0	References.....	4
7.0	Acronyms	4
8.0	Appendices	5
8.1	Appendix A	5

1.0 Executive Summary

Placeholder Text

SUMMARY OF PROGRAM EVALUATION RESULTS	
Strengths	1.
	2.
	3.
Lessons Learned	1.
	2.
	3.
Opportunities	1.
	2.
	3.
Effectiveness Measures	1.
	2.
	3.

2.0 Program Evaluation Approach

[Text here]

3.0 Program Summary

SUMMARY OF PROGRAM EVALUATION RESULTS		
Need	1.	
	2.	
	3.	
Effectiveness	Structure	1.
		2.
	Process	1.
		2.
	Outcome	1.
		2.
Scalability	1.	
	2.	
	3.	

3.1 PROGRAM NEED

[Text here]

3.2 PROGRAM STRUCTURE

[Text here]

3.3 PROGRAM PROCESS

[Text here]

3.4 PROGRAM OUTCOME

[Text here]

3.5 PROGRAM SCALABILITY

[Text here]

4.0 Program Evaluation Results

[Text here]

4.1 STRENGTHS

SUMMARY OF PROGRAM EVALUATION RESULTS	
Strengths	1.
	2.
	3.

Strength #1

[Text here]

Strength #2

[Text here]

4.2 LESSONS LEARNED

SUMMARY OF PROGRAM EVALUATION RESULTS	
Lessons Learned	1.
	2.
	3.

Lesson Learned #1

[Text here]

Lesson Learned #2

[Text here]

4.3 OPPORTUNITIES

SUMMARY OF PROGRAM EVALUATION RESULTS	
Opportunities	1.
	2.
	3.

Opportunity #1

[Text here]

Opportunity #2

[Text here]

4.4 EFFECTIVENESS MEASURES

SUMMARY OF PROGRAM EVALUATION RESULTS	
Effectiveness Measures	1.
	2.
	3.

Effectiveness Measure #1

[Text here]

Effectiveness Measure #2

[Text here]

5.0 Summary

[Text here]

6.0 References

Figure Citations

ITEM	TITLE	SOURCE
[Figure #]		
[Figure #]		

Meetings and Site Visits

DATE	PROJECT SITE	LOCATION	CITY, STATE

Meetings

DATE	MEETING

Program Literature

APA Format

Other Resources

APA Format

7.0 Acronyms

ACRONYM	FORMAL NAME

8.0 Appendices

8.1 APPENDIX A

TEMPLATE O. EVALUATION REPORT FEEDBACK REQUEST

*This email is used when requesting initial feedback from individuals and groups identified in the **Stakeholder Analysis** once the draft **Program Evaluation Report** is complete.*

Dear [Individual or Group Identified in Stakeholder Analysis (Rank and/or Title and Full Name)],

We drafted a program evaluation report for the [name of program], and we would like to request your review and written comments. Please provide comments to the report in a separate document by [date of 30 calendar days of sending the email] at [time (time zone)]. This document is currently in draft and is not approved for distribution.

Thank you for your support.

If you have any question on the report or the process, please do not hesitate to contact me.

Very Respectfully,

[Sender's Signature Block]

TEMPLATE P. EVALUATION REPORT APPROVAL TRACKING FORM

This purpose of this worksheet is to assist program evaluators document and track the approvals needed to finalize the **Program Evaluation Report**.

APPROVER NAME, OFFICE	LAST CONTACT	STATUS
LEADERSHIP		
Name, Office		
STRATEGIC COMMUNICATIONS		
Name, Office		
STAKEHOLDERS		
Name, Office		
OTHERS		
Name, Office		

TEMPLATE Q. FINAL EVALUATION REPORT DISTRIBUTION LETTER

*This email is used to distribute the final **Program Evaluation Report** to interested parties identified in the **Stakeholder Analysis** after obtaining approval from appropriate individuals and offices identified in the **Report Approvals Worksheet**.*

Program Evaluation [DD MM YYYY]

ProgramName
Final Report Communication

Dear [Individual or Group Identified in Stakeholder Analysis (Rank and/or Title and Full Name)],

Attached is the final program evaluation report for the [name of program] program. This report has been reviewed and approved by [noteworthy approvers and strategic communications officers], and can be distributed to [level of distribution allowed (publicly, privately among certain stakeholders)].

We want to thank you again for support provided throughout this program evaluation process. There is no further action required on your part. We hope you will keep us updated on any activities related to [name of program], as well as any noteworthy use or dissemination of the evaluation report.

If you have any questions about the evaluation report or process, please do not hesitate to contact me.

Very Respectfully,
[Sender's Signature Block]

Results Implementation Worksheet

Step

Step 7: Feedback

Purpose

This worksheet will guide translation of evaluation results into lasting program improvements.

Instructions:

This Results Implementation Worksheet helps evaluation teams brainstorm, prioritize and plan follow-up actions stemming from the results detailed in the Program Evaluation Report, and in accordance with program goals and objectives. The template is not a detailed process improvement implementation plan, but rather a structured means to build on the momentum of the program evaluation to plan improvements. The template provides excess space in order to facilitate brainstorming.

The first set of columns in the Implementation Worksheet provide space to itemize the findings from the Evaluation Report

In the next set of columns, the template provides space for the evaluation team to determine first whether action is necessary for a given evaluation result. Potential actions include continuing (with no change), expanding, changing or stopping a certain aspect of the program. Steps needed to execute the change can be added in the next column.

Lastly, the final set of columns prompt the evaluation team to review considerations to effectively implement each action, such as anticipated timeframe for completion, resources required to execute, and stakeholders impacted or whose support could be helpful. Referring back to the Stakeholder Template can assist with identification of appropriate stakeholders.

Results Implementation Worksheet [Template R]

Evaluation Results		Actions		Implementation Considerations		
Result	Result Title	Action Type	Determine the Action Steps	Anticipated Timeframe	Key Resources Needed to Perform Action	Stakeholders
		FINDINGS	Result #1	Continue	Action #1	Days, months, years
Result #2	Expand		Action #2			
Result #3	Stop		Action #3			
Result #4	Change		Action #4			

APPENDIX XI: FREQUENTLY ASKED QUESTIONS

1. Q: Why should our program conduct an evaluation?

A: Program evaluations serve in determining how well a program is working by measuring the achievement of the program's objectives with regard to aspects of program performance. Evaluating results will enable the program to further improve its performance and demonstrate impact, providing insight into the program's potential overall success and ability to endure. A program evaluation can help determine "what works" and "what does not work," showcase a program's effectiveness while establishing it as a "best practice," and improve the quality of services provided to participants. The ability to demonstrate that the program is effective is all the more important during times of fiscal constraint, and having evidence on hand that the program is meeting its objectives may help determine levels of continued funding.

2. Q: What is the difference between Program Evaluation and Performance Improvement?

A: PI is a systematic framework for measuring specific internal processes within an organization. While program evaluation focuses on a holistic overview of an entire program with the goal of determining if the program is reaching its intended goals, PI efforts are more focused, with an assessment of specific processes within a program. PI can also be more flexible compared to program evaluation or research, with the ability to add or drop measures as needed; whereas both research and program evaluation efforts require a formalized data collection process for a set period of time. Another difference is that since the aim of PI is to improve the workings of some aspect of the program: the results of PI efforts are typically only shared with the program staff or program owners. Despite the differences, there is clearly an overlap between the two activities in that they both seek to improve a program's functioning.

3. Q: When does program evaluation cross the line and become research?

A: There is no absolute answer to this question, but generally speaking, if the planned evaluation involves creating groups by randomization that will be considered research. Also, comparison groups are specifically created for the purpose of the evaluation, the effort will likely be deemed research. If the primary purpose of the program evaluation is to publish in the scientific literature, versus to demonstrate that the program is effective for stakeholders, this will likely be considered research. If there are any doubts about whether the program evaluation would be considered research, contacting the IRB point of contact for guidance is recommended.

4. Q: How can I reduce staff resistance to implementing program evaluation?

A: It is not uncommon for a program evaluation to encounter resistance at various levels; Effective leadership strategies exist that can help avoid or mitigate this resistance. Support for the evaluation should be sought early in the process. Leaders should ensure that the staff understands the objectives and value of the evaluation. This support, combined with clear efforts to remain within the scope of the evaluation plan, helps convey legitimacy to the effort. Staff roles should also be delineated from the onset in order to avoid confusion and

guide expectations. Similarly, ensure that a channel for anonymously expressing concerns, providing updates, and disseminating findings exists throughout the process.

5. Q: Where did the DCoE PE Framework come from?

A: In creating this guide, DCoE reviewed several publically available frameworks looking at common steps in the program evaluation process. DCoE’s program evaluation framework closely aligns to other industry standard program evaluation frameworks, even though many different frameworks refer to the same step using different terminology. A table showing how the DCoE framework aligns to other popular frameworks is available in **Figure 433. DCoE Program Evaluation Framework** below.

Figure 433. DCoE Program Evaluation Framework

Publically Available Frameworks	Phases and steps for each framework							
	Planning / Preparation				Execution		Reporting	
<i>Defense Centers Of Excellence In PH TBI</i>	1) Describe program	2) Formulate Evaluation Questions	3) Develop Study Design	4) Develop Data Plan	5) Gather Data	6) Analyze Data	7) Develop Report	8) Act on Findings
<i>Centers for Disease Control</i>	1) Engage stakeholders 2) Describe the program	3) Focus the evaluation design	4) Gather Credible Evidence			5) Justify conclusions	6) Ensure use and share lessons learned	
<i>W.K. Kellogg Foundation Evaluation Handbook</i>	1) Identifying Stakeholders and Establishing an Evaluation Team	2) Developing Evaluation Questions 3) Budgeting for an Evaluation* 4) Selecting an Evaluator*	5) Determining Data Collection Methods		6) Collecting Data	7) Analyzing and Interpreting Data	8) Communicating Findings and Insights	9) Utilizing the Process and Results of Evaluation
<i>NIH- NIDA: Program Evaluation Guide for SUD Treatment Centers</i>	1) Describing your drug treatment program	2) Framing evaluation Questions 3) Refine Evaluation Questions and Select Measures	4) Chose an Evaluation Design	5) Selecting, adapting and developing data collection instruments and procedures	6) Completing evaluation procedures and implementing data collection	7) Analyzing Your Data	8) Reporting and using the findings	
<i>The National Technical Support Center (NTSC) and America's Career Resource Network</i>	1) Define the purpose and scope of the evaluation	2) Specify the evaluation questions	3) Determine the evaluation design	4) Create a data collection action plan	5) Collect data	6) Analyze data	7) Document findings 8) Disseminate information	9) Use the results for program improvement
<i>Public Health Agency of Canada</i>	1) Focus the evaluation (Logic model, stakeholders, Evaluation questions)	2) Select methods (data plan and logistics plan/check) 3) Develop or modify data collection tools			4) Gather and analyze data		5) Make decisions (action plan and report)	

6. Q: How do I know if this Guide is right for my type of program?

A: The Guide is a general framework that can be applied to a broad range of programs. Although many of the examples apply specifically to the MHS, the general methodology can be applied to nearly any program evaluation. It is important to consider which type of program evaluation is needed when applying the Guide to a specific program. As described

throughout the Guide, there are various types of evaluations that may apply to a program, depending on its current stage. Remember, a program is constantly evolving and changing, which means that the type of evaluation conducted must match its current needs to help the program achieve its goals.

7. Q: Who initiates a program evaluation?

A: Program and portfolio managers, as well as other stakeholders and decision-makers, may initiate program evaluations for various reasons. Portfolio leaders may request evaluations to better understand the effectiveness of individual programs with the aim of balancing or growing the portfolio. Program leaders may initiate evaluations to highlight successes and to drive self-improvement, while senior MHS and VHA leaders may request evaluations to inform policymaking, budgeting and decision making. Leaders may also initiate evaluations to provide DCoE with documentation of innovations and best practices for dissemination across the MHS.

8. Q: How long does it take to conduct a program evaluation?

A: The amount of time required to complete a program evaluation depends on the complexity and type of evaluation. Process evaluations that have access to data already being collected may be completed quickly. Conversely, an outcome evaluation that is collecting data on measures of effectiveness over time may be conducted over a period of years. Additional factors are the size of the pool of participants and the length of the intervention. Shorter interventions, such as a one week intensive day treatment program will require less time compared to longer programs like year-long aftercare programs. The number of participants also factors into the timeframe needed. For example even if a program's duration is only two weeks, if it only has the capacity to run eight people through at a time, it might take many months to generate a meaningful sample size.

9. Q: Should we use an external evaluator or do the program evaluation internally?

A: The decision to use an internal or external evaluator is dependent upon the needs of each individual evaluation. Although internally conducting evaluations is often easier, there are circumstances where an independent evaluator will best serve the needs of an organization's program evaluation. External evaluators contribute an outside perspective and objectivity to the evaluation team, along with extensive expertise in program evaluation, statistical analysis and access to a wide range of resources (given their affiliation with external organizations). On the other hand, an internal evaluator has an intimate familiarity with the project's goals and stakeholder community, as well as access to internal resources. Another possibility is to conduct the evaluation internally, but with the technical expertise of an external consultant. These benefits of each perspective should be measured against the priorities of the evaluation.

10. Q: What kinds of skill sets do we need for conducting a program evaluation?

A: Team members with a variety of skill sets are needed to complete a program evaluation. Each program evaluation is unique and will require teams with different skills or experience based on the type of evaluation being performed and the objectives of the evaluation. Another determinant to consider is what stakeholder groups the evaluator will be working

with most often. These questions will largely drive the type of skills the evaluator and support team will require.

11. Q: How do I choose the right data collection tools?

A: In order to select the most appropriate data collection tool or instruments, decision-makers should consider cost, time to administer, specialized training needs, reliability and validity, and requirements/guidance from higher headquarters or governing agencies. While many evaluators wish to include many measures for their evaluation, it is advisable to also consider how having too many questionnaires and other measures affect staff morale and participants' willingness to complete large packets of information.

12. Q: What if I don't have a budget set aside for program evaluation?

A: Unfortunately, planning for the costs of programs often does not include the expense of program evaluation. Even if the program does not have a specific section of its budget set aside for a formal program evaluation, it is still possible to run a smaller scale evaluation using the program's current staff and resources. Many of the designs for program evaluation can be accomplished with little cost and only part time support from staff members. For example, if a substance use disorder (SUD) treatment center is likely already giving several scales upon intake, which could be incorporated into post treatment and follow-up assessments while patients are in aftercare. With the prospect of budget cuts and the increasing number of programs related to psychological health and TBI, it would be advisable to undertake some effort to establish that the program is effective.

13. Q: How do I let others know about the success of my program?

A: Disseminating the results is one of the final important phases of conducting a program evaluation. Several options exist for conveying information, such as reports, brochures, presentations, press releases and so on. The means chosen for delivering this information will be based on the target audience and the specific message to be conveyed. This Guide recommends creating a formal Program Evaluation Report to document and distribute the evaluation results. Note that in some cases it may be necessary to produce multiple versions of a report in order to address the divergent interests of different stakeholder groups. Before releasing the findings of the program evaluation, it is wise to check with the institutions' legal and or public affairs offices.

14. Q: What if the program evaluation shows that my program isn't working?

A: If the program evaluation reveals that the program is not meeting its mission and achieving the objectives and goals it set out to achieve, there are a few options. The program evaluation might reveal key information that will help with process improvement and realigning to the mission. Alternatively, it may be necessary to do an Impact or Cost-Benefit analysis to decide if the program is worth the continued effort and cost or whether it should be discontinued. In most cases, unless the original program evaluation was initiated by key stakeholders who wanted an evaluation conducted to help determine where to cut resources, program evaluation findings can be directed towards process improvement and determining how the program can better achieve its desired outcomes.

15. Q: Where can I get more information about program evaluation?

A: The suggested reading and bibliography at the end of each section of the Guide lists several resources that can contribute to one's knowledge and understanding of program evaluation.

In addition, DCoE can be contacted at:

2345 Crystal Drive,
Crystal Park 4, Suite 120
Arlington, Virginia 22202
(800) 510-7897
Resources@DCoEOutreach.Org

1335 East West Highway
9th Floor, Suite 640
Silver Spring, Maryland 20910
301-295-3257

APPENDIX XII: GLOSSARY OF TERMS

Term	Definition
Activities	As part of a logic model, activities are what the program does with its inputs or resources.
Analysis of Variance (ANOVA)	A statistical technique used to test for differences between multiple (two or more) groups, variables or points in time.
Archival Data	Data that is captured through existing forms, records, policies, etc.
Attrition	A threat to validity; a lack of clarity in results that occurs because participants from one group drop out of a study at a higher rate than participants in other groups.
Case Studies	A method in which a person, event, decision, period, project, policy, institution or system is studied and analyzed for example or learning purposes.
Chi-Square	A statistical technique that is used to compare the goodness of fit between a theoretical frequency distribution and the frequency distribution of observed data.
Cluster Sampling	A method in which the population is divided into groups based on natural commonalities that represent a subset of the population and members are randomly selected from that group.
Coding	The process of examining the raw qualitative data (in the form of words, phrases, sentences or paragraphs) and assigning categories.
Comparison Group	A group of people with similar characteristics and demographics to the treatment group; allows for measurement against the treatment group. Unlike a control group, participants are not randomly assigned.
Control Group	A group of randomly-selected people who do not receive an intervention and are assessed before and after an intervention for comparison with a group that received treatment. The control group should be similar to the experimental group on key variables.
Convenience Sampling	A method that samples individuals who are most convenient and accessible to the researcher.
Correlation	Also referred to as the Pearson Product Moment Correlation, a correlation determines the direction (positive or negative) and magnitude of the relationship between two variables.
Cost-Benefit Analysis	Relates the costs of a program with monetary value of program benefits.
Cost-Effectiveness Analysis	Relates the costs of a program with key outcomes or program benefits.
Data Monitoring Plan	Identifies the process by which data is collected, entered, validated and stored by key personnel. The Data Monitoring Plan should address procedures for collecting, reviewing and safeguarding all data collected.
DCoE Program Evaluation Framework	A program evaluation framework that consists of eight steps, which include: reviewing program information (Step 1); developing evaluation questions (Step 2); designing the evaluation (Step 3); developing the data plan (Step 4); gathering data (Step 5); analyzing data (Step 6), including tests for statistical significance and integrating feedback into the program; developing a report (Step 7); and acting on the findings (Step 8).
Descriptive Statistics	A series of techniques that provide a summary or “big picture” view of the data. Descriptives are a useful first step in data analysis because they can identify overarching themes or trends in the data that help guide later analyses.
Execution Phase	The second phase of the DCoE Program Evaluation Framework. Consists of two steps (Gather Data and Analyze Data) during which the design and plans are put into action, employing the evaluation design to carry out the evaluation, collecting relevant data and examining the data gathered to synthesize findings to report results.

Term	Definition
Experimental Design	Designs used to test the relationship between two variables, specifically whether an independent variable which is manipulated has a measurable effect on a dependent variable. This type of study aims to determine cause and effect relationships. In the case of program evaluations, the interventions or activities a program produces (e.g., therapy sessions, medications, classes) are the independent variables and the dependent variables are the expected outcomes (e.g., reduction in symptoms, return to duty) of the program.
Experimental Group	The group that receives the treatment or intervention under study.
Feedback	The last phase of the DCoE Program Evaluation Framework. Consists of two steps, (Reporting and Acting on findings). Evaluators produce a report of the findings suitable for sharing with stakeholder groups, and evaluation results are applied to improve program services.
Fidelity Assessments	Determine whether a program's activities are being implemented in a manner consistent with the original intention. Fidelity assessments allow program managers to have more confidence that the changes in outcomes are actually due to the interventions (e.g., a PTSD treatment program using prolonged exposure therapy conducts a review of how closely providers are following the treatment protocols).
Flat-File database	A database that uses a single table as its basis. A common means of creating a flat-file database is a spreadsheet program, such as Microsoft Excel. Flat-file databases are simple to use, but lack the flexibility of relational databases.
Focus Groups	A method of qualitative data collection in which a group of people are asked about their perceptions, opinions, beliefs, or attitudes towards a product, service, concept, etc. in an interactive group setting where participants are free to talk with other group members.
Follow-Up Evaluation	An evaluation that queries participants that are no longer actively involved with the program in order to assess the true impact and effectiveness of a program.
Formal Research (FR)	An essential component in the identification of best practices and evidence-based treatment approaches. FR projects require complex methodologies and statistical tools, allow the effects of extraneous variables to be controlled, and offer the greatest level of assurance regarding the validity of the intervention's effects.
Formative Evaluation	Evaluates the quality and delivery of a program to determine whether the program is meeting best practice standards, and an assessment of personnel, procedures, resources, etc., thereby identifying recommendations for restructuring and formation.
Frequency	A measure that describes the number of cases or values within a category.
Goals	Broad, overarching statements of aspiration describing what a program intends to accomplish. They create a context for objectives to break goals down into measurable components. Program goals should always be consistent with the strategic goals, mission and vision of an organization.
Hawthorne Effect	A threat to validity, the Hawthorne Effect refers to a lack of clarity in results that occurs because of participants' awareness of being in a study, a factor that can affect behavior.
History	A threat to validity, history refers to a lack of clarity in results due to an event that occurs between the pretest and posttest of a study (other than the intervention) that may affect the dependent variable.
Impact Evaluation	Evaluations that determine whether the outcomes from a program can be attributed to a particular intervention as opposed to other internal and external influences. Impact analyses usually involve randomized control trails.
Inferential Statistics	Statistics that provide a means to interpret the data and draw conclusions that might be generalizable to larger populations. Provides the ability to look for differences between groups and make judgments regarding the probability that an observed difference between groups is a dependable one rather than one that occurred by chance.
Inputs	As part of a logic model, a program's inputs or resources reflect the resources utilized by a program. In other words, inputs are put IN a program. Resources can be financial, physical or human and can include funding, staff, volunteers or equipment.

Term	Definition
Institutional Review Boards (IRB)	An institutional review board is an oversight body responsible for ensuring that studies or projects that involve human subjects follow necessary ethical standards pertaining to the informed consent of participants and avoidance of unnecessary risk.
Instrumentation	A threat to validity, instrumentation refers to a lack of clarity in results due to inaccurate measures or procedures used in measuring results.
Internal Validity	In a program evaluation, internal validity addresses whether there is a causal relationship between a given intervention and some measured outcome.
Interval data	Data that can be classified and ranked or ordered. It provides meaningful differences between values and is always numerical.
Interviews	A method of qualitative data collection, consisting of a conversation between an investigator and an informant directed towards obtaining information relevant to a specific evaluation. Allows for observation of the respondents' gestures, facial expressions and pauses, and the environment. Can be done face-to-face or over the telephone. An interview can be either structured or unstructured, meaning it can either follow a specific agenda or not.
Judgment Sampling	Selection of a sample population believed to be characteristic of the target population based on the expertise of the researcher.
Logic Model	A graphic that details how a program is structured and how it intends to achieve expected results. Representing a linear sequence, a logic model includes three key components; inputs, outputs and outcomes.
Long-Term Outcomes	Larger community-based impacts of a program, evident several years after program activities.
Maturation	A threat to validity; a lack of clarity in results due to changes that occur in subjects (not due to the intervention) simply due to the passage of time.
Mean	The average of multiple scores; calculated by summing the values and dividing by the total number of scores. The mean can only be calculated for interval or ratio data given the mathematical calculations necessary to derive it.
Measure	A form of measurement based on concrete or objective attributes; a numerical value assigned to an attribute according to defined criteria.
Measures of Central Tendency	Measures which calculate a single value that attempts to describe a set of data by identifying the central position within that data set. The three most common measures of central tendency are the mode, median and mean scores.
Variability	A measure that describes how widely a set of data varies around a particular measure of central tendency, most often the mean. Measures of variability are generally only applicable to interval and ratio data.
Median	The score that separates the upper half of a group of data points from the lower half. The median requires ranking of data before the middle score can be identified; thus it can only be used with ordinal, interval or ratio data.
Metric	A set of measures taken over time that can be presented in the form of a table or graph to identify trends.
Mid-Term Outcomes	Outcomes that result from a program and are evident within a year or two following program activities.
Mission Statement	A broad statement of the program's reason for existence, which describes the program in terms of the program's purpose and philosophy.
Mode	The score that occurs most frequently in a data set; it can be determined regardless of the level of measurement.
Needs Assessment	An analysis that determines what a particular need is, who's need it is, how great the need is, and what might work to meet the need.

Term	Definition
Nominal data	Nominal data is a set of data that uses simple integers as a measurement scale to identify categorical values
Non-Equivalent Comparison Group Design	A type of quasi-experimental design whereby program participants are non-randomly divided into an intervention and a control group. Prior to providing an intervention, both groups are given a pre-test (O1), which is followed by, the intervention group being given an intervention (X), while the control group does not receive the intervention. Once the program intervention is complete, both groups complete a post-test (O2).
Non-Probability Samples	Samples wherein members are selected based on judgment or convenience, and not through a statistical technique. This means that results cannot be generalized.
Objectives	Statements that operationally define the overarching goals of a program by breaking them down into smaller, measurable items. Multiple objectives may exist for each program goal.
One Group, Post-Test Only Design	(Or One-Shot Design); a type of quasi-experimental evaluation whereby participants are all placed into one intervention group. There is no randomized assignment of participants into groups, and no control group; therefor meaningful comparisons cannot be made. The intervention group is given an intervention (X), and at a future point in time data about the intervention effects are collected using a post-test (O).
One Group, Pre-Test/Post-Test Design	A quasi-experimental evaluation where all participants are non-randomly assigned into one intervention group, and there is no control group. The primary distinction from the One Group, Post-Test Only Design is that a pre-test (O1) is given before the program services begin to establish a baseline before the intervention (X) is provided. A post-test (O2) is administered to gather data about the results of the program after the intervention period. This design documents changes that patients have experienced between the time of the pre-test and post-test.
Ordinal	Discrete data which generally refers to categories. It can be classified and ranked or ordered. This type of data generally cannot be analyzed using mathematical operations.
Outcome Evaluation	Evaluations that assess the effectiveness of program implementation with respect to the intended short-term, mid-term and long-term outcomes.
Outcomes	In the context of program evaluation, outcomes reflect the impact of a program and include short and medium term, such as knowledge, skills and individual behavior, to long term, such as social and environmental impacts.
Outputs	Outputs capture the products of a program's activities; they are the units of service.
Performance Improvement (PI)	A systematic framework for measuring specific internal processes within an organization. PI centers on the scope and complexity of focused efforts or processes within a program. PI is aimed at improving the workings of some aspect of a program and the results of PI efforts are typically only shared with the program staff or owners.
Personally Identifiable Information (PII)	Information about an individual that is maintained by an agency and may include their name, address, photograph or information about personal characteristics.
Portfolio	A collection of programs within an organization that address a particular issue; a group of related projects aimed at a particular goal.
Preparation Phase	This initial phase of the DCoE Program Evaluation Framework includes four steps, 1) Review of Program Information, 2) Development of Evaluation Questions, 3) Development of Evaluation Design and 4) Development of a Data Plan. Completing this phase provides insight into the program background and current operations, generates buy-in from stakeholders, formulates evaluation questions, and develops an evaluation design and data plan relevant to the program's goals and objectives.
Probability Samples	A sampling strategy through which every person in a population has a known chance of being selected.
Process Evaluation	Evaluations which identify the need for a program, its target audience, core components, requirements for execution, standardized curriculum and metrics for success.

Term	Definition
Program Evaluation	Individual systematic studies conducted periodically or on an ad hoc basis to assess how well a program is working. The process involves the collection, analysis, and interpretation of data collected to determine the effectiveness of a program's implementation or outcomes and identification of areas in need of performance improvement.
Program Reports	A document prepared to report a specific issue regarding a program, which may provide useful data for an analysis.
Qualitative Analysis	Descriptive or characteristic data, most often collected through techniques such as interviews, observations, journaling, etc. It is difficult to quantify or simplify these data to a numerical value because this type of data often addresses abstract concepts such as emotions, opinions or experiences. As such, qualitative data.
Quantitative Data and Analysis	Analyses most often associated with statistical analyses because of the numerical or empirical nature of the data. Quantitative data quantify something: number of appointments, ratings on a scale of one to five, total score on a measure of anxiety, etc.
Quasi-Experimental Design	A design which seeks to establish the existence of a cause and effect relationship between an intervention and an outcome. The primary difference between a true experimental design and a quasi-experimental design is randomization of group assignment. Quasi-experimental designs are used when randomization of subjects is not possible.
Quota Sampling	A type of sampling that represents a certain characteristic within a population to an extent desired by the researcher.
Range	The most simplistic measure of variability; calculated by subtracting the minimum score from the maximum score of a dataset.
Ratio Data	Ratio data can be classified and ranked or ordered. It provides meaningful differences between data with a zero value. This type of data is always numerical and allows ratios to be calculated between data points.
Regression Analysis	Regression analyses examine the relationship between multiple variables, usually interval or ratio data. Regressions provide information about the effects of changing one variable when the others are held constant.
Rehabilitation	The process of seeking to restore an individual's health or well-being
Reintegration	The re-assimilation of service members, veterans and their families back into society upon return from deployment.
Relational database	A database that incorporates multiple data tables and has established procedures for the tables to merge or interact. These may allow for the sharing of data across networks, via the internet, etc.
Reliability	Consistency of results; notably, that a test yields consistent results.
Root Cause Analysis (RCA)	An analysis performed to determine the cause of part, system and component failure in a program.
Screening	A method used to examine for the presence of any health or psychological condition.
Selection Bias	A threat to validity; lack of clarity in results that occurs due to differences in the experimental group and the control group that existed prior to the start of the study.
Self-Report Measures	A method of data collection in which a person reports or rates his or her feelings, attitudes or beliefs on a topic; may be in various formats. Respondents read the question and select a response without interference.
Short-Term Outcomes	Outcomes expected as a result of and directly following program activities.
Simple Random Sampling	A method of sampling through which any member in the population has an equal chance of being included.
SMART Framework	The SMART Framework assures that units (such as objectives) are SMART: Specific, Measurable, Achievable, Relevant and Time-bound.

Term	Definition
Snowball Sampling	A sampling approach through which initial members are chosen and additional members are identified by original members.
Stakeholder	A stakeholder is any person or group who has an interest in, or is impacted by, the evaluation or the evaluation results.
Stakeholder Analysis	An exercise that consists of devising a strategy of how best to engage each stakeholder. This provides the evaluation team with an opportunity to identify those stakeholder groups and set the tone for a collaborative relationship with these groups.
Stakeholder Engagement Presentation	A tool that provides an opportunity to both inform stakeholders about the evaluation and to gain support for the evaluation. The presentation introduces the rationale for conducting the evaluation, describes the program evaluation methodology, solicits stakeholder input, and provides an opportunity for the team to ask background questions and request program documents to prepare for the remaining steps.
Standard Deviation	A measure of variance or a dispersion of scores from the average.
Standard Operating Procedures (SOP)	A framework to guide a process (for example, it could guide data management staff through required processes, including security measures, for data collection and management).
Statistical Regression	A threat to validity; statistical regression (regression towards the mean) refers to a lack of clarity in results when participants who were selected had extreme scores, because they can be expected to naturally score closer to the mean upon retesting. Therefore, this can give the false impression that an intervention was responsible for change, when in actuality the normalization occurs naturally.
Stratified Sampling	A strategy through which subsamples are randomly drawn from samples of different categories.
Substance Use Disorder (SUD)	Term used to refer to substance abuse or dependence. Symptoms may include tolerance, compulsive behavior and withdrawal. Abuse or dependence may result due to stress, or in combination with PTSD, depression, or other medical conditions.
Suicide	The act of deliberately taking one's own life.
Surveys / Questionnaires	A method of data collection that implements written documents consisting of various questions or scales aimed at gathering information.
Systematic Sampling	A method of sampling through which every "n"th member on a list is selected.
Testing	A threat to validity; refers to lack of clarity in results due to participants learning how to perform better on a test due to experience taking the test.
Threats to Internal Validity	Areas of weakness in an evaluation design that compromise the ability to determine the specific causes of any outcomes measured.
Treatment	A means of managing a physical or psychological condition by providing a prescribed intervention.
t-Test	A tool used to test for differences between two groups, variables or points in time. Depending on the specific characteristics of the data, a single-sample, repeated measures, or independent group t-test may be required.
Validity	Accuracy of content; that a test is measuring what it intends to measure; the relationship between a result and the measure of its true score.

