



Evaluation Approaches for Mental Health Prevention and Early Intervention Programs

Nicole K. Eberhart, Jennifer L. Cerully, Amy L. Shearer, Sandra H. Berry,
M. Audrey Burnam, Patricia A. Ebener

For more information on this publication, visit www.rand.org/t/RR1882

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2017 RAND Corporation

RAND® is a registered trademark.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND

Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

Preface

The RAND Corporation has been conducting an evaluation of the California Mental Health Services Authority (CalMHSA) statewide prevention and early intervention (PEI) programs. In doing so, RAND took a unified approach to evaluating a variety of diverse programs by focusing on core program activities, such as informational resources, training and educational programs, social marketing and media campaigns, and crisis intervention helplines. Furthermore, RAND provided technical assistance to California's counties to support the implementation and evaluation of their PEI activities. In evaluating mental health PEI programs, RAND found itself often breaking new ground: Although there is substantial literature available on evaluating mental health treatment programs, there is very little literature available on evaluating mental health PEI programs in particular. Thus, the RAND team evaluating CalMHSA programs sought to create a guide for those interested in evaluating mental health PEI programs.

RAND Health

This work was conducted in RAND Health, a division of the RAND Corporation, and funded by CalMHSA. A profile of RAND Health, abstracts of its publications, and ordering information can be found at www.rand.org/health.

CalMHSA

The California Mental Health Services Authority (CalMHSA) is an organization of county governments working to improve mental health outcomes for individuals, families, and communities. Prevention and early intervention programs implemented by CalMHSA are funded by counties through the voter-approved Mental Health Services Act (Prop. 63). Prop. 63 provides the funding and framework needed to expand mental health services to previously underserved populations and all of California's diverse communities.

Contents

Preface	iii
Acknowledgments	ix
Introduction and Overview	1
Is This Handbook Right for Your Program? How and Why Was It Developed? What Can You Expect to Find in It? How Do You Use It?.....	1
Where Do I Begin?	1
What Can I Expect to Find?.....	3
How Do I Use the Handbook?.....	4
References	4
CHAPTER ONE	
Basic Steps in Choosing an Evaluation Design	5
Sketch Out a Program Logic Model	5
Identify Core PEI Activities	8
Identify Evaluation Questions.....	9
Select an Evaluation Design	10
Develop a Plan to Collect Data	12
Conduct Basic Data Analyses.....	14
References	19
CHAPTER TWO	
Evaluating Training and Educational Interventions	21
Description of Activity	21
Logic Model	21
Key Evaluation Questions.....	22
Evaluation Design.....	24
Evaluation Data Collection Plan	27
References	29
CHAPTER THREE	
Development and Dissemination of Informational Resources	31
Description of Activity	31
Logic Model	31
Key Evaluation Questions.....	32
Evaluation Design.....	33

Evaluation Data Collection Plan	35
Structure	35
CHAPTER FOUR	
Helpline Operations	37
Description of Activity	37
Logic Model	37
Key Evaluation Questions.....	38
Evaluation Design.....	40
Evaluation Data Collection Plan.....	43
References.....	45
CHAPTER FIVE	
Networks and Collaborations	47
Description of Activity	47
Logic Model	47
Key Evaluation Questions.....	48
Evaluation Design.....	49
Evaluation Data Collection Plan.....	50
References.....	52
CHAPTER SIX	
Social Marketing Campaigns	53
Description of Activity	53
Logic Model	53
Key Evaluation Questions.....	53
Evaluation Design.....	55
Evaluation Data Collection Plan.....	57
References.....	58
CHAPTER SEVEN	
Screening and Referral Services	59
Description of Activity	59
Logic Model	59
Key Evaluation Questions.....	59
Evaluation Design.....	61
Evaluation Data Collection Plan.....	63
References.....	64
CHAPTER EIGHT	
Counseling and Support	65
Description of Activity	65
Logic Model	65
Key Evaluation Questions.....	65
Evaluation Design.....	67
Data Collection Plan	69
References.....	71

CHAPTER NINE

Early Intervention Clinical Services 73

Description of Activity 73

Logic Model 73

Key Evaluation Questions..... 73

Evaluation Design..... 75

Evaluation Data Collection Plan 76

References..... 78

CHAPTER TEN

Conclusion 81

Acknowledgments

This handbook was developed to support high-quality evaluation of mental health prevention and early intervention programs in California and elsewhere. The need for this handbook became apparent when we presented Training, Technical Assistance, and Capacity Building (TTACB) workshops on evaluation design as part of a program implemented by CalMHSA, funded by counties through the voter-approved California Mental Health Services Act (Prop. 63).

The content of this handbook is based on the evaluation design work of a large team, and we are grateful for the team leaders who laid the groundwork for it: Rebecca Collins, Rajeev Ramchand, and Bradley Stein from the RAND Corporation and Michelle Woodbridge from SRI International. Attendees at TTACB workshops also provided valuable information about mental health programs and evaluation methods used in California's counties. In addition, members of the Statewide Evaluation Experts (SEE) Team, a diverse group of California stakeholders, provided valuable input on evaluation design.

The RAND Health quality assurance process employs peer reviewers. This report benefited from the rigorous technical reviews of Joshua Breslau and Donna Farley, which served to improve the quality of this report.

Introduction and Overview

Is This Handbook Right for Your Program? How and Why Was It Developed? What Can You Expect to Find in It? How Do You Use It?

The goal of this handbook is to provide guidance to practitioners and organizations interested in evaluating mental health prevention and early intervention (PEI) programs. In the course of developing an evaluation plan for a series of PEI programs implemented statewide by the California Mental Health Services Authority (CalMHSA), we found many resources that discussed general approaches to evaluating mental health programs and policies but few that discussed how to tailor an evaluation approach to mental health PEI activities in particular. PEI activities may require different evaluation strategies than treatment interventions do. First, the activities being administered are different in that they are not clinical interventions. Indeed, PEI activities are highly diverse and provided in many different types of nonclinical settings. Thus, the process outcomes that are used as indicators of successful implementation tend to look different for PEI interventions. Second, the ultimate targets of change may be different; while treatment interventions generally seek to improve symptomatology or functioning of individuals, that is sometimes a bridge too far for PEI programs, which more commonly seek to improve individual knowledge or skills, and the ultimate long-term mental health outcomes of interest may be at the level of a population rather than the individual.

This handbook addresses the need for a discussion specific to PEI activities: The handbook will be a valuable resource for readers wishing to begin or improve an evaluation of any type of mental health PEI activity. This handbook is intended for readers who already have a basic understanding of program evaluation principles.¹

Where Do I Begin?

It is helpful to begin by considering the goals of your program and the activities you are engaging in to work toward those goals. Think about your PEI program as having two aspects: strategic and tactical components. The *strategic* component is the overall program, which has overarching goals. Those administering the program then make decisions about which activities are needed to achieve these goals, and the individual initiatives that the program chooses

¹ For readers interested in familiarizing themselves with evaluation tools for other prevention programs, we recommend the Getting To Outcomes technical report (Chinman, Imm, and Wandersman, 2004), as well as the program-specific manuals available on the Getting To Outcomes web page (RAND Corporation, undated).

and carries out are the *tactical* components. Progress toward overarching goals is a function of both (1) the mix of activities chosen and (2) the effectiveness with which these activities are implemented. This handbook focuses largely on how to evaluate the tactical components of a program, but it should be pointed out that if the right tactical components are not chosen, then the strategic goals of the program will not be realized.

For example, a program may have an overall goal of preventing suicides (strategic component). This program may seek to achieve that goal with multiple activities, including a social marketing campaign focused on increasing awareness and a training program focused on increasing confidence to intervene with those at risk for suicide. An evaluation of these tactical components would focus on whether they are successfully implemented and whether they achieve their goals related to awareness and confidence to intervene. However, if these are not the right activities in the first place, the program will not achieve its strategic goal of preventing suicides.

Therefore, in approaching PEI program evaluation, the first step is to identify the overall strategic goals of the program that are the driving force behind all the program's efforts. The next step is to choose the mix of activities that will enable the program to progress toward achieving those goals. The scientific research literature should be used to identify activities with the greatest chance of effecting desired changes. Then, the evaluation should address both the strategic design of the overall program and the tactical initiatives being implemented. However, this handbook mainly concentrates on the level of the tactical initiatives. Although this handbook largely focuses on evaluating how well tactical activities are performing, there should also be a broader evaluation of how each tactical component fits within and contributes to the overall strategic goals of the program. Again, progress toward strategic goals could be due to either the strategic design of the program (i.e., choice of activities to pursue) or the effectiveness of the activities as implemented, or both.

To evaluate the strategic components of the program, the program should identify and answer key questions about the overall performance of the program. Some questions to consider are

- How well is the program moving toward its goals?
- How well is each initiative performing?
- What initiatives are most important or least important to the overall program?
- What initiatives are not part of the program but should be added?
- What are the implications of evaluation findings for specific tactical components for potential modifications of the overall program design?
- What are the implications of specific evaluation findings for improving the effectiveness of individual components of the program that may not be performing as needed?

These questions are most effective when they are open-ended rather than close-ended (yes or no), although the answers to the questions may be either qualitative (descriptive) or quantitative (numerical), depending on the question.

The program evaluation could be conducted by the program itself or by a third party that is not involved in program design or implementation. An external evaluation may be ideal because objectivity is ensured. However, self-evaluation may be more cost-effective, and ongoing self-evaluation facilitates quality improvements. Self-evaluation can also be useful and appropriate if there are safeguards in place to preserve objectivity. An external review panel or

advisory committee would ensure objectivity but may not be an affordable option. Internally, adoption of standards-based evaluation with predetermined standards is helpful in ensuring objectivity.

What Can I Expect to Find?

In Chapter One, we sketch out the basic steps involved in choosing and customizing an evaluation plan for mental health PEI activities, largely focusing on the strategic component of the program but also touching on tactical components. These steps include developing an evaluation logic model, identifying core PEI activities, identifying evaluation questions, selecting an evaluation design, and developing a data collection plan. While this handbook focuses primarily on evaluation design—rather than implementation—we do include a brief discussion of conducting basic analyses of data that might be collected in the course of an evaluation.

In the following chapters, we discuss how best to apply these evaluation design steps to eight diverse types of PEI activities (i.e., the tactical components of the program):

1. **training and educational interventions**, including training various types of gatekeepers (such as school personnel or health care providers), training speakers who make presentations to community audiences, directly delivering educational presentations to community audiences, and training specialized staff (such as helpline workers or media producers)
2. **development and dissemination of informational resources**, including the creation of informational resources (e.g., fact sheets, resource guides, online resources) for broad audiences
3. **operation of mental health helplines**, including increasing the availability of helplines and improving the quality of existing lines
4. **networks and collaborations** among organizations, such as a network of helplines or colleges working toward shared goals
5. **social marketing campaigns**, including the use of marketing techniques, such as websites and radio or television advertisements, to promote messages to targeted audiences related to prevention and early intervention
6. **screening and referral services**, such as mobile screening programs (such as health fair screenings); peer-advocate programs; and integrated primary care, mental health, or alcohol- and drug-screening programs
7. **counseling and support services**, including clinical counseling services provided at community wellness centers and family resource centers; parenting, home visitation, and family-preservation programs; school-based anti-bullying and violence-prevention programs; positive youth development programs; and peer-to-peer support, senior advocacy, and socialization programs
8. **early intervention clinical services**, including early intervention clinical services for individuals at risk of depression, anxiety, trauma, or early onset psychosis.

We chose these eight activities because they were pertinent to our work on the evaluation plan for CalMHSA's statewide PEI programs.

How Do I Use the Handbook?

Not all activities covered in this handbook will be relevant to all readers—we expect that the chapter on training and educational interventions will have the broadest relevance—and some PEI activities, such as parent training or outreach to at-risk youth, are not discussed here. However, the basic approach to designing an evaluation for PEI programs that we outline here (see Chapter One) will be useful to all readers, regardless of what specific PEI activities they are interested in evaluating. We recommend that most users of this handbook read Chapter One’s overview of PEI program evaluation design and then the chapters on the specific activities that are similar to the activities the users are evaluating: Chapter Two for training and educational interventions, Chapter Three for informational resources, Chapter Four for mental health helplines, Chapter Five for networks and collaborations, Chapter Six for social marketing campaigns, Chapter Seven for screening and referral services, Chapter Eight for counseling and support services, and Chapter Nine for early intervention clinical services. We also encourage readers to explore chapters on activities that may not be a part of their PEI programs, because they provide illustrative examples of how to apply and make best use of the basic evaluation design principles discussed here.

References

- Chinman, Matthew, Pamela Imm, and Abraham Wandersman, *Getting To Outcomes™ 2004: Promoting Accountability Through Methods and Tools for Planning, Implementation, and Evaluation*, Santa Monica, Calif.: RAND Corporation, TR-101-CDC, 2004. As of December 12, 2016:
http://www.rand.org/pubs/technical_reports/TR101.html
- RAND Corporation, “Getting To Outcomes®: Improving Community-Based Prevention,” web page, undated. As of April 24, 2017:
<http://www.rand.org/health/projects/getting-to-outcomes.html>

Basic Steps in Choosing an Evaluation Design

Good structure increases the likelihood of good process, and good process increases the likelihood of a good outcome.

—*Avedis Donabedian (1988)*

Sketch Out a Program Logic Model

Often, the best way to begin planning your evaluation is by creating a logic model. A logic model, or logic framework, is a graphical depiction of the relationship between program inputs, activities, and outcomes. Logic models help you take stock of the various resources and activities that make up your program, as well as the expected program outcomes. The simple process of listing out each of these elements can be surprisingly helpful in clarifying your evaluation needs and goals.

The logic model can be done at two different levels: the overall strategic level (overarching program goals and the activities selected to advance those goals) and the tactical level (specific activities that are implemented). In this chapter, we give an example of a *program* logic model (see Figure 1.1), which is at the higher strategic level. At this level, it is important to first consider the overall program goals. Describe what this program is trying to accomplish. Then these goals (i.e., outcomes in the logic model) should guide the development of the program's design and implementation (i.e., structures and processes below), which should be informed by the research literature.

At the tactical level, PEI programs often consist of multiple PEI *activities*, such as the activities that are the subject of each of the following chapters in this handbook. For each of these broad activities, you will want to create separate logic models (examples of specific activity logic models are provided in each chapter).

Logic models are useful tools both for program planning (e.g., to help understand what needs to be done to get to outcomes) and for evaluating why and how well programs are working. This handbook focuses largely on the utility of logic models in terms of program evaluation.

Of the various approaches to constructing logic models, we find the structure-process-outcome model to be a simple and helpful one. This model allows for an understanding of the different kinds of activities being implemented and how they affect outcomes. The structure, process, and outcome boxes of the logic model can be linked to distinct evaluation questions about structure, process, and outcomes, which we discuss later in the chapter.

What Is a Structure?

Structures are the elements of your organization and specific PEI program that affect how activities are conducted. Structures are resources that facilitate program implementation, such as the building and physical equipment, staffing, method of service delivery (e.g., over-the-phone versus in-person screening), informational resources (e.g., training manuals), and organizational networks. Structures include existing PEI capacities and resources, as well as those that you are developing, including networks, training curricula, educational resources, marketing materials, policies, and protocols. Structures are usually easy to identify and measure because they are readily observable.

In your strategic *program* logic model, structures could include the mix of initiatives, who is carrying them out, and the overall administrative structure to oversee the work. Your list of structures may include all the physical resources available to your program (which may consist of numerous PEI activities). In your tactical PEI *activity* logic models, you need only include those relevant to the specific PEI activity (e.g., in your helpline logic model, you need only include resources used in implementing this activity), but you would include additional details of how that activity is organized.

What Is a Process?

Processes refer to what is actually done by the program—that is, the intervention or activities that are conducted. Processes are the mechanisms through which structures lead to outcomes. For example, for a training curriculum (a structure) to have an impact, trainings need to be provided and people need to participate in them; therefore, *participation in training* is a process outcome. Other examples of processes are provision of outreach and social marketing efforts (process outcomes could be the number of people viewing a billboard, receiving an event flyer, or listening to a radio ad) or provision of helpline services (with the number of calls taken as a process outcome).

In your strategic *program* logic model, your process list should include broader PEI activities (e.g., training and technical assistance, dissemination of informational materials). In your tactical PEI *activity* logic model, your process list should be specific to each PEI activity (e.g., in your training and technical assistance activity logic model, processes would only include detail about the training and technical assistance activities). In addition to quantitatively tracking implementation (e.g., number of people reached), an evaluation should qualitatively (i.e., descriptively) examine how well implementation is going and why it is or is not going well.

What Is an Outcome?

Simply put, **outcomes** are the effects of the program—for example, the changes in attitudes, behavior, knowledge, or individual functioning that can be attributed to program activities. As with structures and processes, these outcomes can occur at both the strategic (program) level and the tactical (activity) level of a specific activity. For example, a program may have an overarching strategic goal of preventing suicide, as well as more-specific goals at the tactical level (such as improved media reporting of suicide or increased knowledge of how to intervene with those at risk for suicide) that are designed to bring it closer to the strategic goal.

While the structure and process evaluation may help you better understand *why* the objectives are or are not being met, stakeholders are often primarily interested in knowing

whether the program is meeting its objectives. Outcomes are often time-limited (that is, expected within a certain period) and involve a specific change in a target population.

Program effects are most often evaluated in terms of the impact of a program on those who are directly exposed to the program, and effects are usually assessed either immediately after exposure or within a period of months after exposure. We refer to these direct program effects as **short-term outcomes**. Short-term outcomes are the *immediate* targets of change, and they are often quantitative (i.e., numerical) measures of progress toward goals.

In contrast, **long-term outcomes** for the program are often qualitative—descriptively, did the program accomplish what it intended to accomplish? These more long-term outcomes are the major motivation for the program. However, they are typically much more difficult to measure in the context of a specific program evaluation, in part because the effects may be indirect and may be influenced by factors outside the program. For example, a teacher who gains knowledge about student mental health issues through a training program may directly or indirectly provide support to a student, and that student then may be less likely to drop out of school in later years. Long-term outcomes are sometimes referred to as *impacts* because they are related to broader challenges facing individuals, families, and society as a whole that can be the result of mental health challenges that PEI programs are generally intended to address.¹ Programs seek to affect long-term outcomes, such as graduation or suicide rates; however, for a number of reasons, including the small number of events and the long time frame that would be needed to assess them, it is inherently difficult to assess these long-term effects within the scope of a program evaluation. Thus, this handbook focuses largely on evaluation of the short-term outcomes that are the direct, immediate effects of the program—effects that are believed to be causally connected to the ultimate, long-term outcomes of interest.

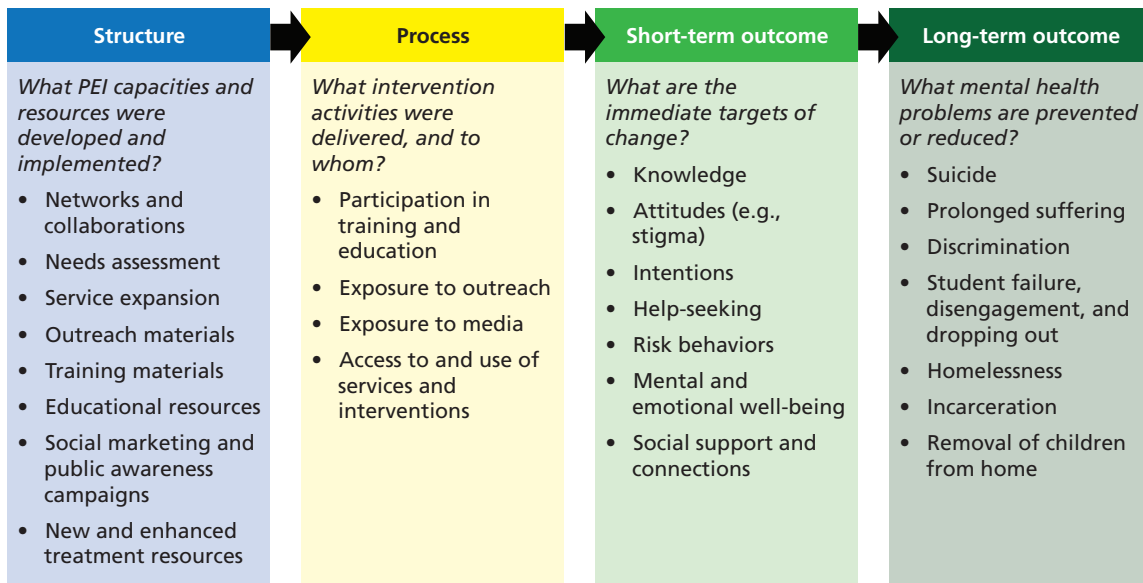
While long-term outcomes are included in logic models and mentioned briefly throughout this handbook, we do not explore in any significant detail how to assess these long-term outcomes. Figure 1.1 is a sample logic model showing various structures, processes, and outcomes that may be applicable to many different PEI interventions. Each of the major components of the logic model is in its own box.

In addition to evaluating outcomes relevant to program goals, it may also be useful for the evaluation to assess unintended outcomes—i.e., unanticipated effects of the program, whether they are positive or negative. If a program results in undesired outcomes, there is likely a need to change the structures and processes—i.e., what is being implemented and how.

Although we explained the elements of a logic model in their logical order, when creating your own program logic model and, more broadly, when approaching program evaluation, it can be useful to start with the outcomes first and work backward. These outcomes should reflect the goals and objectives of the program, which are the targets for the evaluation to examine. In the outcome box of the model, list out the outcomes that your program aims to achieve, being as specific as possible. It can be helpful to look at how the strategic goals of the overall program fit with the goals of the different tactical activities—i.e., if the tactical outcomes are achieved, will this push the program toward the strategic outcomes of interest?

¹ For readers interested in an evaluation strategy that links PEI programs to longer-term impacts, there is a separate RAND report by Watkins and colleagues (Watkins et al., 2012) that recommends that the California state government use and improve existing public health surveillance data to monitor key mental health–related outcomes. This monitoring would inform PEI program planning and provide information to evaluate, over time, the combined effects of the state’s investment in PEI programs. The report provides a conceptual evaluation framework for linking PEI programs to key outcomes and impacts, as well as detailed information regarding existing data sources and priorities for data improvements.

Figure 1.1
Program Logic Model



RAND RR1882-1.1

Next, think about the activities or processes your program is conducting to achieve these outcomes, and list them in the process box. In the structure box, list all of the resources your program uses in its activities (or processes).

There is value in assessing each of the logic model components, along with the links between them. For example, when evaluating a training program, it is helpful to understand the resources and curricula used in implementation (structure), how these structures affect the number of people trained (process), and whether those trained experienced an increase in knowledge as a result of the process (short-term outcome). However, depending on the evaluation goals and the resources available, an evaluation may choose to focus on just one or two major components of the logic model. For example, an evaluation of structures and processes alone would provide valuable information on factors related to program implementation, even if such factors cannot be related to outcomes. For more information on logic models specific to PEI mental health programs, see the RAND report *Evaluating the Impact of Prevention and Early Intervention Activities on the Mental Health of California's Population* (Watkins et al., 2012).

Identify Core PEI Activities

People often approach evaluation with the question, “Did the program work?” However, programs are often composed of multiple and diverse tactical activities that are theoretically linked to the overarching strategic program goal. Therefore, it is more informative to break down the program into component activities and ask, “How did each specific *activity* work?” This makes it possible to pinpoint which specific activity or aspect of a program is driving any program-level effects. Teasing out the individual influence of each program activity can help

you determine which activities are contributing positively to your intended program outcomes, as well as which activities are less effective and in need of revision. This is helpful in making strategic-level decisions about the right mix of tactical activities to include in the program. Lumping together diverse activities, some of which are effective and some of which are not, to ask about *program* effectiveness could mask the positive effects of a particular activity or group of activities and thus lead to inaccurate findings about program-level effects. However, there may be a trade-off regarding how much the program is split up into components and the ability to isolate effects of those specific components.

In the following chapters, we present information about how to evaluate eight categories of PEI activities: training and education of various target populations, including gatekeepers, speakers, school personnel, helpline staff, speakers, and media producers; development and dissemination of informational resources; operation of helplines; networks and collaboration with other organizations; social marketing campaigns; screening and referral services; counseling and support services; and early intervention clinical services. Identifying which PEI activity (or activities) you want to focus your evaluation on will help you identify and refine your tactical-level evaluation questions. These questions about specific activities can be used to inform program-level questions about the appropriate mix of activities to include.

Identify Evaluation Questions

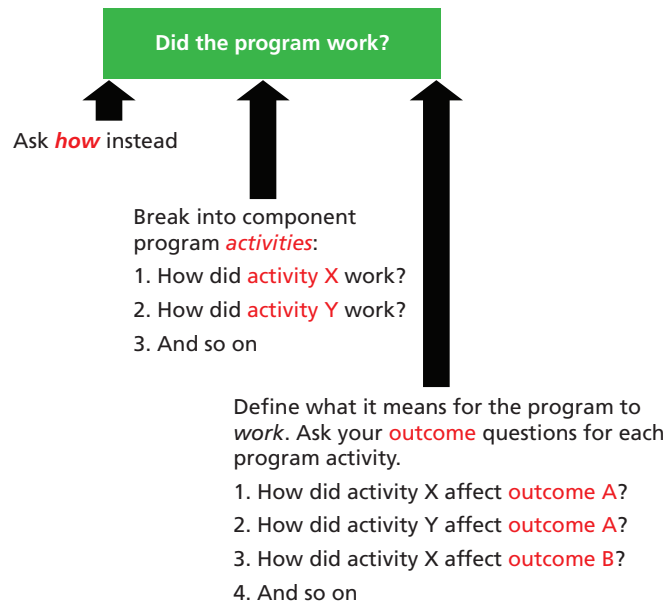
In the section above we suggested asking, “How did each specific *activity* work?” instead of “Did the *program* work?” There are three important reasons to frame the question this way. The first, discussed above, is that we are interested in activity-level effects. Second, asking *how* the activity worked, rather than *whether* it worked, means that your evaluation results will provide you with detailed, useful information about how your activities led to your outcomes—the processes involved.

Third, the simple question “did the program work?” implies that there is one definition for *work*. In reality, your program could be influencing a myriad of short- and long-term outcomes. You defined these outcomes of interest (i.e., what it means for your program to work) when you created your logic model. Your evaluation questions will be designed around them. These steps are illustrated in Figure 1.2.

Lastly, the effects of your program activities could differ by individual factors (such as age, gender, and race) or activity-specific factors (such as level of participation, when the activity occurred, and where it was being implemented). Therefore, evaluation questions should be as specific as possible and take into account not only *how* the activity worked but also for *whom* it worked and *when* and *where* it worked. This means further refining your program activity outcome questions for specific populations and scenarios. Specific examples for each PEI activity are provided in the chapters that follow.

At the *structure* step, it is important to ask questions about the resources or capacity that will be developed to implement a program activity. Returning to the example of teacher training regarding student mental health, a structure question might be, “What is the approach to training?” At the *process* step, it is important to answer questions about the deployment of resources to achieve program goals—for example, “What was the reach or dissemination of a particular activity to its target population?” For example, how many teachers received the training, and how many students can the teachers potentially reach? At the *outcome* step, it is

Figure 1.2
Identifying Evaluation Questions



RAND RR1882-1.2

important to answer questions about the impact of an activity on those in the target population who participated in or were exposed to the activity. For example, did the training result in increased teacher knowledge about how to intervene with at-risk students?

The evaluation questions you ask may be driven by your program's own goals (e.g., quality assurance or improvement) or by outside sources (e.g., funding bodies or program stakeholders).

Select an Evaluation Design

At its simplest, an evaluation design is a plan for what to measure and how to measure it. Examples include the structures, processes, and outcomes of interest; what kinds of measurement tools are most appropriate; which population to measure; and when to measure (are baseline and follow-up measurements needed to assess changes over time?). In this section, we briefly review different kinds of evaluation designs that may be used for different evaluations purposes. First, we describe descriptive, nonexperimental evaluation designs that can be used to meet accountability requirements. Then, we describe some evaluation designs that go beyond accountability to measure an activity's effectiveness. We introduce and describe advantages and disadvantages of the following approaches:

- to meet accountability requirements
 - descriptive
- to measure effectiveness
 - experimental
 - quasi-experimental
 - pre-post.

What Evaluation Design Should Be Used for Ongoing Data Collection to Meet Accountability Requirements and Inform Quality Improvement?

The most basic kind of evaluation design is **descriptive** and nonexperimental. This approach provides basic information on the resources and capacities developed; what intervention activities were implemented, and to whom; and how participants fared with respect to outcomes of interest at a single point in time. This approach is relatively easy and inexpensive to implement, but it does not allow one to draw conclusions about whether the outcomes observed are truly due to specific PEI activities. However, given that it is not overly burdensome to implement, it can be used on an ongoing basis for external reporting and internal continuous quality-improvement purposes.

What Evaluation Design Should Be Used to Examine the Program's Effectiveness?

If the goal of the evaluation is to determine whether the activity affects outcomes, then stronger, and typically more costly, evaluation designs are needed. The strongest but most costly design is an **experimental design**, in which participants are randomly assigned to either a group that receives the intervention or a control group of individuals that does *not* participate in the intervention. The intervention and control groups are compared to determine whether there are differences between the groups because of the program or intervention. Members of control groups are carefully selected to be similar to participants in intervention groups, and this design provides data on the same outcome measures as on program participants. The intervention and control groups are measured at two different points in time, and the changes over time are then compared. This method provides the strongest evidence that an intervention is effective, but it is expensive and can be impractical to administer in real-world settings. Often, it is simply not an option to randomly assign who will receive a mental health PEI intervention and who will not.

Less rigorous, but easier to administer, is a **quasi-experimental** design, in which there is no random assignment to intervention and control groups; rather, information is collected from individuals who received the intervention and individuals in a control group who did not (for example, consumers at a neighboring agency that did not implement a particular PEI program). The quasi-experimental design is considered less rigorous than a true experimental design because there may be preexisting differences between groups that make it more difficult to interpret results. A related disadvantage is that it can be difficult to identify an appropriate control group for a quasi-experimental design.

When it is not possible to identify any comparison group, a **pre-post**, or before-after, design allows for observing changes in outcomes over time. This approach is more rigorous than a purely descriptive design, although it is less conclusive than experimental and quasi-experimental designs. With this approach, it is not possible to determine definitively whether changes observed in participants are due to the intervention—rather than other factors—since there is no control group. Thus, it is important to be careful in interpreting results of pre-post designs, as it is possible that any changes observed could have occurred even without the intervention.

Because community agencies often lack the capacity to implement these more-rigorous approaches involving control groups, most of the evaluation designs discussed in the following sections on specific PEI activities are either descriptive or pre-post designs. However, we mention designs with control groups here because they provide more-robust evidence related to the effectiveness of an intervention. If promising results are obtained with descriptive or pre-post

designs, a next step might be to implement a more rigorous quasi-experimental or experimental design.

Develop a Plan to Collect Data

When implementing a data collection plan, it is helpful to think about data collection for structures, processes, and outcomes.

Structure

Data collection for structures is primarily a matter of systematically identifying the resources that contribute to operating the PEI program being evaluated. Most of the data related to structures will be qualitative (i.e., nonnumerical) in nature. Typical data collection techniques will include document review and key informant or focus group interviews. Key informants are people with the greatest knowledge of the structures developed as part of a PEI program. If the evaluators already have detailed knowledge of the program (i.e., because they are directly involved in administering the program), the same kinds of information typically collected via these formal data collection techniques could be gathered using informal methods.

To conduct a document review, relevant documents must first be collected and stored. Review of the materials should be systematic, with the same key pieces of information obtained from all similar documents. To this end, it is helpful to develop a *data abstraction* tool, such as a spreadsheet in which each document is represented by a row, and each piece of information to be abstracted from the document is represented by a column. The chapters on specific activities provide some helpful suggestions about what kinds of information to obtain from various kinds of documents.

To conduct key informant or focus group interviews, key informants must first be identified. When possible, it is helpful to identify multiple informants to obtain multiple perspectives on the issues being queried and to ensure that important information is not omitted. Informants can be interviewed either individually or as a focus group. Individual interviews may be more time-consuming, but it may be easier to elicit information from people who are reluctant to speak in a group setting. Focus group interviews allow for the group to come to a consensus about important issues, but it can be challenging to ensure that all voices in a group are heard. For information on conducting effective focus groups, see *Focus Groups as Qualitative Research*, published by SAGE Publications (Morgan, 1996).

Process

Data related to processes may be either qualitative or quantitative in nature. Qualitative techniques, such as key informant or focus group interviews, can help highlight issues related to program implementation. Quantitative data collection may also be helpful, depending on the specific activity and the goals of the evaluation. Surveys can facilitate an understanding of the experiences of those receiving, as well as administering, the PEI intervention, and they can also help you better understand the fidelity with which an intervention is administered (i.e., if it was implemented as it was intended to be implemented). A survey can include both open-ended and close-ended questions. Close-ended questions (i.e., questions in which a respondent chooses among discrete response options) provide data that are easier to analyze. In contrast, open-ended questions (i.e., items for which the respondent can write in an answer) provide

richer information but are more complicated to analyze because they require interpretation, often through the development and application of a coding scheme.

These data collection methods can be used to examine process questions about whether the intervention has been implemented as intended and who received the intervention. One way to examine whether the intervention was implemented as intended is for trained raters to observe the administration of the intervention and rate the fidelity. This kind of evaluation requires a rating scheme, and raters have to be trained to conduct ratings that are consistent with each other and the intent of the rating scheme.

Perhaps the most important aspect of processes to assess is the reach of the PEI activity—that is, how widely the activity was administered and which groups participated in the activity. For some kinds of interventions (e.g., those that are considered health care services), data on reach may be readily available via *administrative* records—data routinely collected for administrative or billing purposes. For other kinds of intervention, it may be necessary to collect reach data. Data collection techniques will vary for different aspects of reach. Logs are helpful in understanding the number of times a PEI activity was implemented. Sign-in sheets or preregistration forms are helpful in understanding the number and characteristics (e.g., demographics) of the recipients of the intervention activity. If it is not feasible to include demographic information on registration materials or sign-in sheets, surveys may be used to identify the number of people in key demographic groups that were reached.

Throughout this handbook, we provide suggestions for activity-specific process measures, but an excellent place to start is the process measure directory available from the Center for Quality Assessment in Mental Health's National Inventory of Mental Health Quality Measures database (Center for Quality Assessment and Improvement in Mental Health, undated).

Outcome

Data related to outcomes will be primarily quantitative and should be obtained either through surveys or existing records. The first step is to determine what kinds of outcomes are of interest (i.e., what the PEI program is intended to accomplish). Broadly, many PEI programs seek to effect changes in knowledge, attitudes, skills, and behaviors. In Table 1.1, we present outcomes that could be measured for three common types of mental health PEI programs—programs aimed at reducing stigma and discrimination, preventing suicide, and improving student mental health.

Some of these outcomes can be measured using existing records, such as administrative records, medical records, school records, and vital statistics. Others will require new data collection, via surveys. Although new outcome measures can be developed for an evaluation, it is often preferable to use existing measures. Using existing measures is a more efficient use of resources than developing new ones, it leverages work others have already done to ensure that the measure is reliable and valid, and it enables easier comparison to other programs evaluated with the same measures. To find appropriate measures, it can be helpful to consult databases of psychological measures or search the scientific literature for research studies that have measured the same outcomes of interest. Table 1.2 includes a list of possible resources for identifying appropriate outcome measures; this list is not comprehensive but is intended as a starting point.

When choosing an outcome measure, it is important to look for measures that are **reliable** (i.e., likely to be used by participants the same way each time they complete it), **valid** (i.e., actually measures the concept it aims to measure), short (to minimize burden on respondents), free (publicly available and not requiring a copyright fee), easy to understand, and easily translated into different languages if needed. Sometimes there is no existing mea-

Table 1.1
Possible Outcomes to Measure

Stigma- and Discrimination-Reduction Resources	Suicide Prevention Resources	Student Mental Health Resources
<ul style="list-style-type: none"> • Knowledge and recognition of common mental health disorders • Beliefs about recovery and the effectiveness of mental health treatment • Attitudes toward individuals with mental illness • Desired social distance from individuals with mental illness • Behavioral intentions toward individuals with mental illness • Behavior toward individuals with mental illness 	<ul style="list-style-type: none"> • Knowledge about suicide risk factors and resources for at-risk individuals • Attitudes toward at-risk individuals • Behavioral intentions to refer at-risk individuals to mental health resources • Referrals of at-risk individuals 	<ul style="list-style-type: none"> • Teachers' knowledge about risk factors for students and resources • Students' knowledge of mental health resources • Teacher attitudes toward and comfort with making referrals • Teacher referrals of at-risk students

sure for the outcome of interest, in which case new items can be developed. However, interpretation of these items will be limited if there is no evidence for their reliability or validity.

Conduct Basic Data Analyses

This section is intended to serve as an introduction to conducting basic analyses of the types of data you may collect during an evaluation. Again, this is not a comprehensive guide to statistical analyses but rather a starting point for making sense of the data that an organization collects. We describe two types of data—qualitative and quantitative—and describe some basic approaches to analyzing each. These suggestions are summarized in Table 1.3.

Qualitative Data

What Are Qualitative Data?

Qualitative data are nonnumeric and descriptive, such as those obtained from the following sources:

- responses to open-ended survey questions (i.e., questions that ask for a narrative response rather than an endorsement of a discrete category or provision of a numerical response)
- key documents or materials (e.g., program curricula, fact sheets, memoranda of understanding [MOUs])
- key informant interview and focus group transcripts
- transcripts from chat or text conversations with a helpline
- social marketing messages.

Basic Approaches to Analyzing Qualitative Data

Two basic approaches to analyzing qualitative data include careful review and narrative summary of available data. Qualitative data sources should be reviewed for relevant types of information, which can then be tracked in a separate data collection instrument. For example, you might assess the structure of a training program by reviewing training curriculum documents and conducting key informant interviews about the approach of each training program and

Table 1.2
Some Possible Resources for Identifying Appropriate Outcome Measures

Resource	Description of Resource	Location of Resource
RAND literature review of suicide prevention program evaluations	A literature review of suicide prevention programs that includes a discussion of outcome measures used to evaluate them	Acosta et al., 2012 (www.rand.org/t/TR1317)
RAND literature review of mental health stigma- and discrimination-reduction program evaluations	A literature review of mental health stigma- and discrimination-reduction programs that includes a discussion of outcome measures used to evaluate them	Collins et al., 2012 (www.rand.org/t/TR1318)
RAND literature review of student mental health program evaluations	A literature review of student mental health programs that includes a discussion of the outcome measures used to evaluate the programs	Stein et al., 2012 (www.rand.org/t/TR1319)
Community Tool Box	Resources for developing, implementing, and evaluating community interventions created by the Work Group for Community Health and Development at the University of Kansas	Work Group for Community Health and Development, undated (www.ctb.ku.edu/en)
Substance Abuse and Mental Health Services Administration (SAMHSA) National Registry of Evidence-Based Programs and Practices	A registry of evidence-based mental health and substance abuse intervention programs, including information on the quality of outcome measures used to assess each intervention	SAMHSA, undated (www.nrepp.samhsa.gov/)
National Consortium on Stigma and Empowerment Resources	A list of resources available on the National Consortium on Stigma and Empowerment website, including a tool kit containing measures	National Consortium on Stigma and Empowerment, "Resources," web page, undated (www.stigmaandempowerment.org/resources)
Social-Personality Questionnaire Instrument Compendium	A website that links to various social and personality psychology measures	Alan Reifman, Social-Personality Questionnaire Instrument Compendium, last updated March 8, 2017 (www.webpages.ttu.edu/areifman/qic.htm)
PsycTESTS®	A research database of psychological tests maintained by the American Psychological Association (subscription required)	American Psychological Association, PsycTESTS®, database, undated (http://www.apa.org/pubs/databases/psyc-tests/index.aspx)
RAND Online Measure Repository	An online searchable database containing measures related to psychological health and the treatment of traumatic brain injuries that was created to support program monitoring and evaluation	RAND Corporation, undated (www.rand.org/nsrd/ndri/centers/frp/innovative-practices/measure.html)

Table 1.3
Data Types and Recommended Basic Analyses

	Qualitative Data	Quantitative Data
Definition	Nonnumeric, descriptive data	Numeric data
Example	Responses to open-ended survey questions, interview transcripts	Responses to close-ended survey questions, characteristics of audiences reached
Recommended basic analyses	Identify topics, concepts, or themes of interest; review data for these dimensions; and create a descriptive summary.	For categorical data or data with two possible outcomes, generate frequencies or percentages. For continuous data, calculate means and standard deviations. Use t-tests to compare means across time and between groups of interest where possible.

the degree to which existing scientific evidence was used to guide program development. Information about the dimensions of interest (in this example, training approach and evidence base for the training) can either be extracted verbatim from the qualitative data sources or paraphrased in a data collection spreadsheet (see Table 1.4 for a simple example). The spreadsheet can include as many rows as needed to describe data from all sources, and it can include as many columns as necessary to track data needed to inform the evaluation of structures, processes, and outcomes.

Analysis of this data entails reviewing the spreadsheet and summarizing it in narrative form. Harking back to the example above, the spreadsheet could be reviewed to describe the variety of training approaches used and the proportion of training programs that are evidence-based (i.e., they use content and communication strategies supported by scientific evidence). If it is of interest, each individual program can be described in detail. Organization of the narrative summary will depend heavily on the purposes of the evaluation and the intended audience.

A full review of qualitative data collection and analysis is beyond the scope of this handbook. Many practical resources are available online, such as those listed in Table 1.5. For a more in-depth discussion of a common analytic approach, see Braun and Clarke (2006).

Quantitative Data

What Are Quantitative Data?

Quantitative data are any data in numeric form, such as those obtained from the following sources:

Table 1.4
Example of Using a Spreadsheet to Track Qualitative Data About Program Structures

Data Source Reviewed	What Is the Training Approach?	Is the Training Evidence-Based?
Transcript from key informant interview about Program A	Interviewees watch a one-hour educational lecture about the consequences of mental health stigma.	Research suggests that providing educational presentations about mental health literacy and consequences of stigma and discrimination can reduce stigma (e.g., Penn and Couture, 2002).
Program B curriculum document	Attendees view a 30-minute presentation by a person with mental health challenges and then form small groups to discuss their impressions.	Research suggests that contact with a person with mental health challenges is an effective way to reduce stigma (Couture and Penn, 2003).

- responses to close-ended survey questions
- data describing how many people used a service (e.g., number of people trained, number of people in an audience, number of calls or callers to a helpline, number of people who used a mental health service), and their demographic characteristics
- data tracking communications between organizations
- data tracking dissemination of materials.

Basic Approaches to Analyzing Quantitative Data

To conduct simple analyses of quantitative data, it is often helpful to run basic descriptive statistics, such as frequencies, percentages, or means (i.e., averages). A frequency is the number of times an outcome or response occurs—for example, the number of survey respondents out of the total sample (e.g., 54 out of 150) who indicate that they have sought mental health treatment in the past 12 months. Frequencies are best used to assess data that fall into categories (e.g., race, number of people indicating that they accessed a particular website) or that have only two possible outcomes (e.g., yes or no responses, ethnicity [not Hispanic/Latino or Hispanic/Latino]). In lieu of or in addition to frequencies, percentages may also be helpful; for example, when describing audience characteristics, it is often useful to know what percentage of the total audience falls into certain categories (e.g., reporting that 36 percent of survey respondents have sought mental health treatment).

Means (often referred to as *averages*) are informative when trying to understand how to interpret data that are continuous (e.g., age, responses to survey questions that are given on a seven-point scale). For example, suppose that a social marketing campaign about mental health is implemented on a college or university campus. Students may be asked how much they liked a campaign message on a scale ranging from one (“not at all”) to seven (“a lot”). Calculating the mean score for this question across all students answering the survey can provide a sense of how much students like the messaging, on average.

Standard deviations are a measure of the spread or variation in data and can be helpful in interpreting means. For example, consider the previous example of how much students like a social marketing message. Understanding where the average falls on the seven-point scale is useful. However, more meaning can be made of this by exploring whether there is a small or

Table 1.5
Qualitative Evaluation Resources

Name	Description	Location
Centers for Disease Control and Prevention, “Analyzing Qualitative Data for Evaluation”	Brief overview of using qualitative data in evaluation	Centers for Disease Control and Prevention, “Analyzing Qualitative Data for Evaluation,” Evaluation e-TA: Evaluation Briefs, No. 19, April 2009 (www.cdc.gov/healthyyouth/evaluation/pdf/brief19.pdf)
National Science Foundation, <i>User-Friendly Handbook for Mixed Method Evaluations</i>	In-depth chapters on qualitative methods and analysis in evaluation	Joy A. Frechtling and Laure M. Sharp, <i>User-Friendly Handbook for Mixed Method Evaluations</i> , Arlington, Va.: National Science Foundation, Directorate for Education and Human Resources, Division of Research, Evaluation, and Communication, 1997, Chapters 3 and 4 (www.nsf.gov/pubs/1997/nsf97153/start.htm)
The Pell Institute and Pathways to College Evaluation Toolkit	Step-by-step walk-through of collecting and analyzing qualitative evaluation data	Pell Institute and Pathways to College Network, “Analyze Qualitative Data,” The Pell Institute and Pathways to College Evaluation Toolkit, 2017 (http://toolkit.pellinstitute.org/evaluation-guide/analyze/analyze-qualitative-data/)

large standard deviation. If the standard deviation is small, it suggests that most respondents liked the campaign at similar levels. A large standard deviation would indicate that students really varied in whether or not they liked it. This information provides important context and could influence decisions about strategies that could be used to revise the messaging.

Note that in some cases it may be better to calculate the median, or middle value, of a set of data, rather than the mean. Calculating the median is most useful if there are some data points that look very different from the others. Going back to the example of the survey question about how much students like a social marketing message, there are many ways that survey respondents could use the seven-point scale. For example, multiple respondents could provide responses between one and seven (resulting in a large standard deviation), or most respondents could give responses that are close to three or four (resulting in a small standard deviation). Respondents could also give responses that cluster together with a few outliers (i.e., extreme values). For example, most respondents could select a one or two, with just a few respondents responding with a seven. In a case with an outlying data point of seven, the mean will be higher than one or two, even if one or two are the ratings that most respondents gave. When the data are largely clustered together with only a few outliers, it is preferable to calculate the median to get a better understanding of the data, as opposed to the mean.

Basic descriptive statistics can be calculated to analyze information about the groups of most interest. Consider again the example of the campus survey. If administrators were particularly interested in learning how responses to the message varied by students' year in school (e.g., freshman, senior) or race, administrators could cross-tabulate means based on these two characteristics. To create cross tabulations (also referred to as *contingency tables*), descriptive statistics can be calculated for subgroups. For example, mean scores could be calculated for each subgroup created by crossing year in school with respondent race, resulting in a table of mean scores for all students by year in school and race.

Comparing Data Across Different Groups or Points in Time

One of the most powerful ways to make meaning of quantitative data is to compare descriptive statistics, either across groups or across time. To illustrate the way that comparing across groups could be useful, consider the example of an educational program designed to increase mental health literacy. To help determine the effectiveness of the program in increasing knowledge about mental health issues, a survey containing items testing knowledge of mental health issues could be administered to participants after training, as well as to a control group of people similar to the training participants but who did not actually undergo the training. Mean responses on the knowledge survey items can be compared between the group of training participants and the control group. A higher mean score for training participants relative to the control group participants would suggest that the training was effective at increasing mental health knowledge. Similar scores between the groups would suggest that the training did not meet the goal of increasing knowledge. Comparing descriptive statistics across time can also be very informative. For example, this same survey could be administered to participants in the training program, both before and after the training session. Mean scores on the knowledge questions could be calculated. The scores from before the training can be compared with those obtained after to determine training effectiveness. Higher post-training knowledge scores relative to pre-training scores would suggest that training was effective, while similar scores would suggest that the training had no effect.

Visual inspection of means provides some idea of what might be occurring in the data set, but this process is not sufficient because differences in means may occur because of chance. To make sure that means do not just differ by chance, statistical tests must be conducted that take into account the size of the data set and the variation within it when comparing means.

Although the example of evaluating a training program has been used throughout this section, thinking strategically about what comparisons to make and when, can inform any evaluation using quantitative data. For example, if an organization is making targeted efforts to disseminate a social marketing campaign to Hispanics and Latinos, the organization could compare the percentage of Hispanics and Latinos reached at the beginning of the campaign with the percentage reached at the end of the campaign. To strengthen the assessment of the targeted efforts, the organization could also compare the growth in reach among Hispanics and Latinos with the growth in reach among other racial and ethnic groups during the same period.

More-Complex Approaches to Analyzing Data

More-formal methods of analyzing both qualitative and quantitative data exist, each requiring more-complex considerations. For example, qualitative data can be analyzed by developing systematic coding schemes and conducting formal content analysis. Quantitative data can be analyzed using advanced statistical techniques (e.g., regression, analysis of variance [ANOVA]) that provide greater flexibility in answering evaluation questions. For example, more-advanced techniques would be needed to compare data both between different groups and across multiple points in time, as well as to determine the effect of one variable on the data while holding all others constant (e.g., learning how race is related to responses, even when accounting for other factors that might differ, such as age or gender). A good overview of these kinds of statistical analysis techniques is found in *Applied Linear Statistical Models* (Kutner et al., 2013), but if these methods are new to you, it is best to explore whether and how to conduct these types of analyses in partnership with experts in these specific analysis techniques.

References

- Acosta, Joie, Rajeev Ramchand, Lisa H. Jaycox, Amariah Becker, and Nicole K. Eberhart, *Interventions to Prevent Suicide: A Literature Review to Guide Evaluation of California's Mental Health Prevention and Early Intervention Initiative*, Santa Monica, Calif.: RAND Corporation, TR-1317-CMHSA, 2012. As of May 3, 2016: http://www.rand.org/pubs/technical_reports/TR1317.html
- American Psychological Association, PsycTESTS®, database, undated. As of April 25, 2017: <http://www.apa.org/pubs/databases/psyc-tests/index.aspx>
- Braun, Virginia, and Victoria Clarke, "Using Thematic Analysis in Psychology," *Qualitative Research in Psychology*, Vol. 3, No. 2, 2006, pp. 77–101.
- Center for Quality Assessment and Improvement in Mental Health, "National Inventory of Mental Health Quality Measures," database, Center for Quality Assessment and Improvement in Mental Health, undated. As of December 19, 2016: <http://www.cqaimh.org/NIMHQM.htm>
- Centers for Disease Control and Prevention, "Analyzing Qualitative Data for Evaluation," *Evaluation e-TA: Evaluation Briefs*, No. 19, April 2009. As of April 24, 2017: <https://www.cdc.gov/healthyyouth/evaluation/pdf/brief19.pdf>

- Collins, Rebecca L., Eunice C. Wong, Jennifer L. Cerully, Dana Schultz, and Nicole K. Eberhart, *Interventions to Reduce Mental Health Stigma and Discrimination: A Literature Review to Guide Evaluation of California's Mental Health Prevention and Early Intervention Initiative*, Santa Monica, Calif.: RAND Corporation, TR-1318-CMHSA, 2012. As of May 3, 2016:
http://www.rand.org/pubs/technical_reports/TR1318.html
- Couture, Shannon, and David Penn, "Interpersonal Contact and the Stigma of Mental Illness: A Review of the Literature," *Journal of Mental Health*, Vol. 12, No. 3, 2003, pp. 291–305.
- Donabedian, Avedis, "The Quality of Care: How Can It Be Assessed?" *JAMA*, Vol. 260, No. 12, 1988, pp. 1743–1748.
- Frechtling, Joy A., and Laure M. Sharp, *User-Friendly Handbook for Mixed Method Evaluations*, Arlington, Va.: National Science Foundation, Directorate for Education and Human Resources, Division of Research, Evaluation, and Communication, 1997. As of April 25, 2017:
<https://www.nsf.gov/pubs/1997/nsf97153/start.htm>
- Kutner, Michael H., Christopher J. Nachtsheim, John Neter, and William Li, *Applied Linear Statistical Models*, 5th ed., Boston: McGraw Hill, 2013.
- Morgan, David L., *Focus Groups as Qualitative Research*, Thousand Oaks, Calif.: Sage Publications, 1996.
- National Consortium on Stigma and Empowerment, "Resources," web page, undated. As of April 25, 2017:
<http://www.stigmaandempowerment.org/resources>
- Pell Institute and Pathways to College Network, "Analyze Qualitative Data," *The Pell Institute and Pathways to College Evaluation Toolkit*, 2017. As of December 18, 2016:
<http://toolkit.pellinstitute.org/evaluation-guide/analyze/analyze-qualitative-data/>
- Penn, David L., and Shannon M. Couture, "Strategies for Reducing Stigma Toward Persons with Mental Illness," *World Psychiatry*, Vol. 1, No. 1, 2002, pp. 20–21.
- RAND Corporation, RAND Online Measure Repository, database, undated. As of April 25, 2017:
<http://www.rand.org/nsrd/ndri/centers/frp/innovative-practices/measure.html>
- Reifman, Alan, Social-Personality Questionnaire Instrument Compendium, last updated March 18, 2017. As of April 25, 2017:
<http://www.webpages.ttu.edu/areifman/qic.htm>
- SAMHSA—See Substance Abuse and Mental Health Services Administration.
- Stein, Bradley D., Lisa Sontag-Padilla, Karen Chan Osilla, Michelle W. Woodbridge, Courtney Ann Kase, Lisa H. Jaycox, Elizabeth J. D'Amico, Jennifer L. Cerully, Nicole K. Eberhart, and Shari Golan, *Interventions to Improve Student Mental Health: A Literature Review to Guide Evaluation of California's Mental Health Prevention and Early Intervention Initiative*, Santa Monica, Calif.: RAND Corporation, TR-1319-CMHSA, 2012. As of April 24, 2017:
http://www.rand.org/pubs/technical_reports/TR1319.html
- Substance Abuse and Mental Health Services Administration, National Registry of Evidence-Based Programs and Practices, undated. As of April 25, 2017:
<http://www.nrepp.samhsa.gov/landing.aspx>
- Watkins, Katherine E., M. Audrey Burnam, Edward N. Okeke, and Claude Messan Setodji, *Evaluating the Impact of Prevention and Early Intervention Activities on the Mental Health of California's Population*, Santa Monica, Calif.: RAND Corporation, TR-1316-CMHSA, 2012. As of April 24, 2017:
http://www.rand.org/pubs/technical_reports/TR1316.html
- Work Group for Community Health and Development, Community Tool Box, University of Kansas, 2016. As of April 25, 2017:
<http://ctb.ku.edu/en>

Evaluating Training and Educational Interventions

This chapter provides an overview of how to approach evaluation design and implementation for a specific kind of mental health PEI activity, training and educational (TE) interventions, and to provide guidance to those seeking to evaluate this kind of intervention. We begin by describing what kinds of activities are included in this category. We then review what a logic model for TE activities might look like. We then use the logic model framework of structures, processes, and outcomes to describe some key evaluation questions for TE interventions and review what evaluation design might look like for these activities. Finally, we discuss how to approach data collection for TE activities. We follow this basic road map through evaluation design and implementation for all the specific types of evaluation activities discussed in the following chapters.

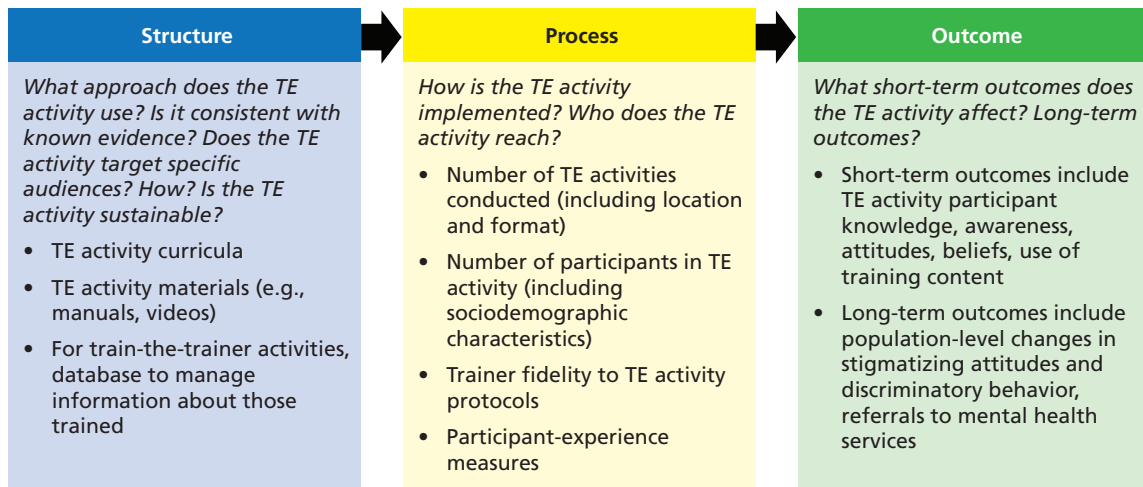
Description of Activity

Many intervention strategies for addressing mental health issues, such as suicide prevention and stigma and discrimination reduction, involve training various types of gatekeepers (e.g., educators or health providers), training speakers who make presentations to community audiences, or directly delivering educational presentations to community audiences. For example, a mental health advocacy organization may offer educational events to the general public on how to recognize and respond appropriately to signs of mental health challenges. Or the organization may offer train-the-trainer programs, in which participants are trained to lead educational or training presentations. TE programs often aim to target a variety of outcomes, such as knowledge and beliefs about, and attitudes and behavior toward, people with mental health challenges.

Logic Model

In planning your evaluation, it is helpful to first sketch out a logic model for your TE intervention (for example, see Figure 2.1). As described in Chapter One, *structures* are the capacities and resources that are available for use by the TE intervention. These can include materials developed as part of the training, such as training curricula or a database of available trainers, as well as meeting rooms, technology, and equipment. *Processes* are actually done by your intervention—the training activities and the people to whom they are delivered. Process information can include who conducted a TE activity and the format of the training or

Figure 2.1
TE Logic Model and Key Evaluation Questions



RAND RR1882-2.1

educational session (e.g., group discussion, webinar). *Outcomes* are the possible effects of the TE activity and can include short-term outcomes, such as increased knowledge and awareness, as well as long-term outcomes, such as lower rates of discriminatory behavior.

Key Evaluation Questions

Key evaluation questions for TE activities may be driven by your program's own goals (e.g., quality improvement) or by outside sources (e.g., funding bodies or program stakeholders). Figure 2.1 lists examples of questions that can guide the evaluation of the structure, process, and outcome of a TE activity. In this section, we discuss the kinds of questions that can be asked and the data that might be collected to answer these questions. The evaluation questions you choose will depend on the type of information you are interested in and your organization's capacity to collect and analyze the relevant data.

Key questions related to evaluating the **structure** of a TE activity include

- **What approach does the TE activity take in training and educating others?** For example, one activity might develop a curriculum focused on training crisis helpline volunteers to appropriately respond to callers, while another might focus on changing public attitudes toward people with mental illness.
- **Is there an evidence base for the design of the TE activity?** For example, does a TE activity that aims to reduce stigma toward people with mental illness encourage trainees to interact with people with mental illness, a strategy for which there is existing evidence of effectiveness (Corrigan et al., 2012)?
- **Does the TE activity target specific audiences? If so, how?** Many activities aim to reach specific audiences, such as racial and ethnic minority groups. Understanding how these groups are targeted (e.g., through TE activities conducted in different languages

or materials adapted to be compatible with the culture of the audience) provides useful information about how the structure of the TE activity helps achieve the goals of the program.

- **Is the TE activity sustainable?** For example, consider whether the resources available to the activity require extensive upkeep (e.g., a database of available trainers), are stable (e.g., printed curricula and materials), or are self-sustaining (e.g., resources or training sessions that bring in income to pay for themselves). The purpose of asking this question is to better understand the factors, such as funding sources and staffing plans, that may enhance a TE activity's longevity. Questions about sustainability should consider not only financial sustainability but also staff and consumer support, interest, and time investments.

To answer key questions about the structure of a TE activity, information can be gleaned from a variety of sources, such as formal documentation of program development or interviews with people who were instrumental in program development. Any materials developed as part of the TE activity (e.g., training curriculum documents and manuals or materials for participants or people leading TE activities) are also rich sources of information about the structure of an activity.

Key questions related to evaluating the **processes** involved in a TE program include

- **What specific activities were carried out to implement the TE intervention?** To answer this question, you will need to collect data on how, when, and where the TE activities are conducted. Examples include: What was presented to those trained? How was it presented? How much was presented (e.g., number and duration of training sessions)? In what settings? How closely did the training adhere to the protocol for the training activity (a concept known as *fidelity*)? What were participants' and trainers' experiences with training (e.g., participant-satisfaction measures)?
- **Who does the TE activity reach?** This question helps better understand exactly who a program affects. For many TE activities, this includes members of the audience. For train-the-trainer programs, it is worth considering assessing secondary reach as well (i.e., the audiences that newly trained trainers are likely to reach).

Addressing process questions requires data from a number of sources. As mentioned above, it may be necessary to collect data about when and where TE program sessions are held and who delivers them, as well as data about the numbers of audience members and their characteristics. Other useful data to collect include participant and presenter experiences of the TE activity (e.g., survey items asking about participant satisfaction and enjoyment of the activity) and presenter fidelity to TE program protocols and curricula. Note that the number of participants in a program and their satisfaction with it are not outcome measures but rather measures of the process by which the program was delivered.

Key questions related to evaluating the **outcomes** of a TE activity include

- **Did the TE activity affect the targeted short-term outcomes?** Short-term outcomes are the immediate changes that a TE activity is designed to bring about. These often include participant knowledge, attitudes, and intentions to behave in a certain way (e.g., to use the content of the training in some way). For example, a TE activity aiming to help sec-

ondary school teachers identify students with potential mental health problems and refer them appropriately may target changing knowledge about mental health symptoms and referral strategies, attitudes about students with mental health problems, and intentions to refer.

- **What are the long-term outcomes?** Long-term outcomes are the ultimate changes, such as preventing suicide or decreasing discrimination toward people with mental illness, that a TE activity is designed to bring about. As noted in Chapter One, these outcomes are often influenced indirectly by the activity or at a future point in time, making them difficult to measure in the context of a specific program evaluation. For this reason, we do not focus on long-term outcomes in the remainder of this chapter.

Addressing short-term outcome questions often requires survey data assessing the outcomes that the activity targets, such as participants' knowledge, awareness, attitudes, beliefs, and intentions to use program content. For train-the-trainer programs, it may be helpful to assess the constructs for both training participants and audience members of the TE presentations conducted by newly trained trainers. Where possible, data about behavior changes of program participants (e.g., the number of referrals made by teachers, evidence of actual policy changes made by activity participants) can be gathered.

Evaluation Design

The design of your evaluation should be structured to answer the evaluation questions you have identified. There are a number of decisions to be made about the design of an evaluation for a TE activity, including what to measure, how to measure it, who will be measured, and when the measurements will occur. Below, we present several design considerations relevant to evaluating structures, processes, and outcomes of TE activities. We are not suggesting that an evaluation design for a specific TE activity should include all of the design elements discussed here; rather, the evaluation design should be specific to the goals of the program, and its evaluation and should incorporate the elements needed to accomplish those goals.

Structure

Answering key evaluation questions about a TE activity's structures requires qualitative methods, such as **document reviews** and **key informant interviews**. A review of existing materials (e.g., formal documentation of program development, developed curricula) can shed light not only on a TE activity's approach and target audience but also on the degree to which it is consistent with existing scientific evidence. Key informant interviews with those who were instrumental in TE program development can also help answer key questions about a TE program's structures.

For example, a program curriculum document can be reviewed to determine the approach of the training and whether a literature review was conducted or experts were consulted as part of the development process. Program materials can be reviewed to determine whether and how they are tailored to meet the interests of the specific target audiences. Depending on the depth of TE activity materials, more information (or supplemental information) may be obtained through key informant interviews focused on the same topics. To address program sustainability, again, materials can be reviewed for factors affecting the ability of the program to con-

tinue (e.g., staffing and other costs, funding streams), and more information about this can be obtained through key informant interviews. For more information on how to analyze program costs and funding streams for sustainability, see Bond, Morrison-Saunders, and Pope (2012).

Process

Answering questions about TE activity processes also entails qualitative and quantitative methods, including key informant interviews, fidelity assessments, participant surveys, and the examination or collection of data on program reach.

Understanding how activities were implemented can be addressed through key informant interviews with individuals who played critical roles in program implementation (who may or may not be the same individuals interviewed about TE activity structures), or this information can be gathered through informal means if the evaluator is also involved in program design or implementation.

The fidelity with which presenters conduct TE activities can be assessed through one of several methods. Presenters can be observed and rated by experts while conducting TE activities, or additional survey items can be added to surveys that ask participants to recall whether the presentation included all program elements. Presenters can also self-report whether they included all elements, but this method may result in biased responding if presenters cannot recall exact details or if they feel pressure to respond positively.

If many TE activities occur, it may be prohibitive to assess the fidelity of all presentations. In this case, selecting a random sample of presentations (e.g., by numbering all presentations that are scheduled to occur in a given month, selecting ten random numbers, and assessing fidelity at only those presentations) for which fidelity assessments would be conducted is a possible solution. Random sampling is important because it means that the presentations assessed do not differ in some systematic way; the results, therefore, are likely to apply to the broader group of presentations. This approach is far preferable to more-biased selection methods, such as choosing presentations given by the suspected strongest presenters or only those given in one region or county. If one of these samples were selected, conclusions could be drawn only about the group the sample represents (for example, conclusions would apply only to the strongest presenters, representing a best-case scenario, or conclusions would apply only to presentations in a specific region and not those in other areas).

For some TE activities, data on program reach may be routinely collected for administrative purposes. For others, it may be necessary to collect new reach data. Learning about the reach of the program entails collecting data on the number and characteristics of participants attending programs. When considering what characteristics to collect as part of this process, it is important to define the key demographic groups that the TE activity targets (e.g., racial and ethnic minorities, individuals who identify as lesbian, gay, bisexual, queer, or transgender [LGBTQ]). See the data collection section below for a more detailed discussion on the collection of data to understand reach.

Surveys administered to participants may also be used to collect data on reach (i.e., participant demographics), as well as to collect process measures, such as satisfaction with TE activity presenters and content.

Outcome

Addressing the short-term outcomes affected by a TE activity requires multiple considerations of what outcomes to measure, when, and from whom.

What Outcome Measures to Assess

Selected outcomes should be those most relevant to the goals of the TE activity. For example, if the main focus of a TE activity is to train police officers to recognize the symptoms of mental illness or to recognize a person in crisis and respond appropriately, then it will be important to assess knowledge of mental illness symptoms and referral processes. Because many programs ultimately hope to change participants' behavior, it is desirable to collect behavioral data where possible (e.g., police department data on the number of times participating officers asked for assistance from responders trained in working with those in crisis). It may also be helpful to collect self-reports of how participants made use of information or strategies presented as part of a TE activity. See the data collection section below for more information on how to measure selected outcome variables.

When to Assess Outcome Measures

After selecting outcomes to measure, the next step is to determine when to measure them. Although measuring outcomes immediately after a TE activity (i.e., postactivity) may seem sufficient, postactivity outcome data are hard to interpret because it is unclear whether there was any change (positive or negative) in responses relative to how participants would have responded prior to participation. Thus, we strongly recommend collecting data from participants prior to activity participation (i.e., baseline data), as well as postactivity. This way, baseline data can be compared with postactivity data to determine not only whether participants' responses changed but also the magnitude of change, as well as whether it was in the expected direction. Many TE activities hope for lasting changes in outcomes (e.g., ongoing positive attitudes toward people with mental health problems), and not just changes immediately after the activity. To assess how long the effects of the TE activity last, it is worth collecting postactivity data immediately after a TE activity and again several weeks or months later.

Who Should Complete Outcome Measures

A helpful first step in deciding who should be sampled is to consider who will provide the most-useful information for determining the outcomes of the TE activity. So, for a train-the-trainer activity, it is worth considering collecting data from activity participants, as well as the people they will go on to conduct presentations for later. However, for a TE activity targeting the general public, collecting data from activity participants is likely sufficient. Depending on the number of TE presentations made, however, it may be prohibitive to collect data at each one. Harking back to the discussion of random sampling when conducting fidelity assessments, we recommend selecting a random sample of presentations from which to collect outcome measures. The reasons are similar: to avoid bias and to be able to draw broader conclusions about the effects of the TE activity.

In addition to sampling considerations, the use of a control group should be considered. A control group is a group of individuals carefully selected to receive the same outcome measures as program participants so that the responses of members of the control group can be compared with the responses of program participants. Ideally, control groups are as similar as possible to groups receiving TE activities. For example, if a TE activity were being delivered to students in several classrooms in a particular school, then a reasonable control group would be the students in another classroom of the same grade level, either at the same school or at another school with similar characteristics. Because factors other than TE activities (e.g., heavy media coverage of an event involving people with mental illness) can influence program par-

ticipants, being able to compare their changes in outcome measures with members of a control group can help to more effectively determine whether changes are due to the TE activity or to the influence of changes in knowledge, attitudes, or beliefs in the broader social environment. Even if baseline and postactivity data were collected and showed improvement, it would be unclear whether this improvement was truly due to the TE activity rather than external factors, unless a control group did *not* show similar improvement.

Evaluation Data Collection Plan

In this section, we outline important considerations for developing an evaluation data collection plan for TE activities.

Structure

In the evaluation design section, we suggested using **document reviews** and **key informant interviews** for addressing key evaluation questions about the structure of a TE activity. To review documents, materials must be collected, stored (online or in hard copy), and made available to the people who will be conducting the evaluation. A data collection instrument will facilitate this process. For example, in an Excel spreadsheet, each row could represent an item being reviewed, and each column could represent content to be abstracted from the material being reviewed, including information about program approach, whether consistency with an evidence base was considered, whether the program targets specific audiences, and how. Information about resources that will persist (e.g., materials made available online, such as a database of those who have completed train-the-trainer programs), program funding, and program staffing can be included in the spreadsheet to inform judgments of sustainability.

To incorporate key informant interviews into an evaluation, you must first identify several key informants and then draft an interview protocol detailing the questions you would like to ask. Interviewers should then be trained in how to conduct interviews (if the interviewers have no, or only minimal, prior experience) and how to record information obtained during the interviews. To collect the data, interviewers may either take notes directly on a data collection form that corresponds with the interview protocol or make an audio or video recording of the interview and transcribe it later for review (either by the interviewer or another party).

To tailor program materials for specific audiences, we suggest using either participant questionnaires after the training or expert review of program content. See the outcome section below for more information about questionnaire data collection.

Process

To assess processes related to TE activities, we suggest conducting key informant interviews, administering participant and presenter surveys, examining presenter fidelity to program curricula, and collecting data about program reach. Key informant interviews, discussed above, could include questions specific to implementation. Surveys administered to participants attending a TE activity could include questions about implementation (e.g., the credibility, likability, and degree of engagement of program presenters). Program presenters can also be surveyed about program materials (e.g., ease of use, perceived effectiveness, strategies for engaging targeted audiences).

Examining presenter fidelity to TE activity curricula may be accomplished through participant surveys issued after a TE activity is completed or by having trained observers view presentations and rate fidelity. Survey items about fidelity could ask participants to indicate how many required elements of a program they recall encountering. If trained observers conduct fidelity assessments, the observers will likely use an existing fidelity assessment guide or will design one for the project.

Program reach can be assessed by tracking the number of TE presentations delivered, the number of participants at each presentation, and the locations of presentations, as well as by documenting pertinent characteristics of the participants, such as age, gender, racial and ethnic background, sexual orientation, and languages spoken. These data can be collected in a variety of ways: Program participants can provide this information when registering in advance of an event or on an attendance sheet on arrival at a program, they can fill out a card with the information and drop it in a box when the program is over, or they can answer questions about these items on a participant survey. For train-the-trainer programs, *secondary reach* (i.e., who the newly trained trainers reach) can be assessed by tracking similar data among those who see presentations given by the newly trained trainers.

Reach data can be collected through web-based tools (e.g., having advance online registration) or using hard copies of materials (e.g., program sign-in sheets). If hard copies are used, the data will need to be either entered directly into a format that will facilitate data analysis (e.g., Excel spreadsheets) or mailed directly to someone with this capability. Also, to maintain participants' privacy, this information may be collected without asking for any identifiable information (e.g., name, address). However, it may be desirable to collect identifiable data so that data from one source can be linked to another (e.g., linking sign-in data with survey data collected later). In this case, care needs to be taken to allow participants some measure of privacy when providing data (e.g., by allowing participants to seal questionnaires in envelopes before returning them to presenters) and when transporting, storing, and reporting data.

These tracking data can be matched against the need for TE activities in various locations or for various demographic groups (e.g., whether trainings occurred in areas or groups with highest suicide rates or in areas identified in needs assessments). These data can also be compared against an organization's own goals (i.e., if an organization had a goal of delivering a specific number of trainings, to specific groups of individuals, or in specific areas).

Outcome

We suggest using participant surveys as the primary method to obtain outcome data. It is important to make sure that the outcome measures included in a survey are those that will help determine whether the goal of a TE activity is met. For example, if a program aims to reduce stigma toward people with mental health problems, then it would be vital to include questions about attitudes toward these individuals. However, questions about unrelated topics, such as knowledge of mental health resources at participants' children's schools, would not be relevant. See Table 1.1 for an overview of concepts that could be measured for three common types of mental health PEI programs—programs aimed at reducing stigma and discrimination, preventing suicide, and improving student mental health. Although new outcome measures can be developed for an evaluation, it is often preferable to use measures that have already been developed and tested for reliability and validity. Table 1.2 provides a list of some possible resources for identifying appropriate measures. When a measure to assess a concept addressed

in a particular TE activity does not exist, using a combination of newly developed measures and existing measures is appropriate.

References

Bond, Alan, Angus Morrison-Saunders, and Jenny Pope, "Sustainability Assessment: The State of the Art," *Impact Assessment and Project Appraisal*, Vol. 30, 2012, pp. 53–62.

Corrigan, Patrick W., Scott B. Morris, Patrick J. Michaels, Jennifer D. Rafacz, and Nicolas Rüsch, "Challenging the Public Stigma of Mental Illness: A Meta-Analysis of Outcome Studies," *Psychiatric Services*, Vol. 63, No. 10, 2012, pp. 963–973.

Development and Dissemination of Informational Resources

This chapter provides an overview of how to approach evaluation design and implementation for activities that focus on the dissemination of informational resources. We describe what kinds of activities are included in this category and review what a logic model for these dissemination activities might look like. We then use the logic model framework to describe some key evaluation questions for resource dissemination efforts and review what evaluation design might look like for these activities. Finally, we discuss how to approach data collection for PEI activities that aim to disseminate informational resources.

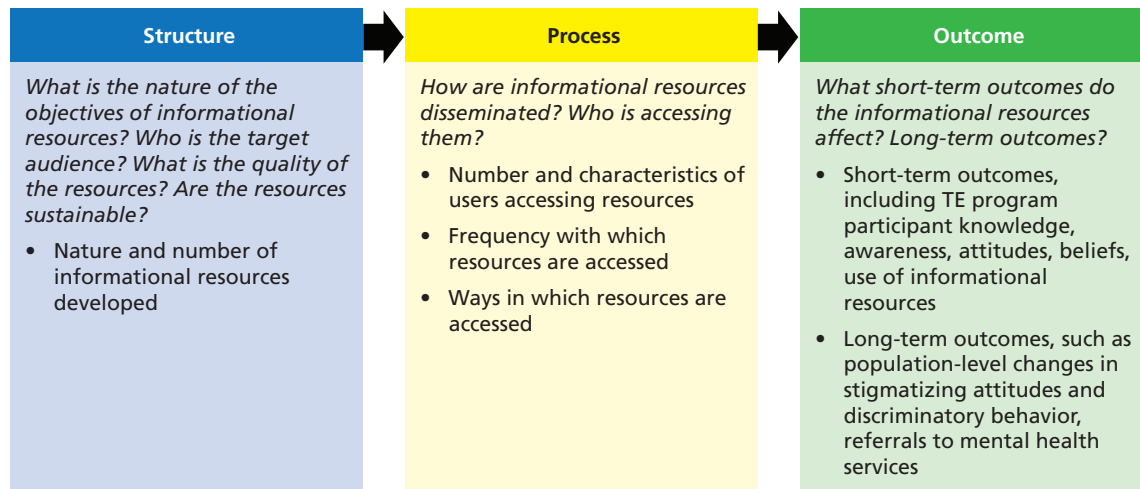
Description of Activity

PEI strategies for preventing suicide, reducing mental health stigma and discrimination, and improving student mental health often involve the creation of informational resources for broad audiences. For example, a university system may develop a set of fact sheets or resource guides to be shared across its campuses; alternatively, an advocacy group may create and disseminate fact sheets to inform people with mental health challenges about laws and policies protecting them from discrimination. These resources are commonly available via websites. For example, for the University of California's Red Folder Initiative, each campus creates an online quick reference guide to mental health resources for faculty and staff to use in assisting distressed students. Informational resources can have broad-ranging topics and goals, making it necessary to think critically about how best to assess their impact, as described below.

Logic Model

Figure 3.1 presents a logic model to guide the evaluation of informational resources. For this model, structures include the nature and quantity of informational resources made available. Processes—the ways in which informational resources are disseminated and used—can include information about how many users acquired the resources and through what methods (e.g., downloading them online versus receiving hard copies at an event). Outcomes are the effects of informational resources and can include short-term outcomes, such as increased knowledge and awareness, as well as long-term outcomes, such as improved student mental health.

Figure 3.1
Informational Resources Logic Model and Key Evaluation Questions



RAND RR1882-3.1

Key Evaluation Questions

Figure 3.1 also includes a list of key questions for evaluating the structure, process, and outcome related to informational resources. We summarize each key evaluation question below and provide some insight into what data are necessary to answer the evaluation questions.

Key questions related to evaluating the **structure** of informational resources include

- **What is the nature of the informational resources?** Answering this question involves assessing the topics that are covered by the informational resources and the objectives of the resources. It also helpful to learn more about the format of the resources and whether they are hosted online, are text- or video-based, and so on.
- **Who is the target audience for the informational resources?** This question addresses details about the target audience, including what strategies were used to tailor resources for the target audience. For example, a tool kit developed for schools may involve resources developed both for teachers and for students and would likely be tailored differently for these two audiences (e.g., different content and reading levels).
- **What is the quality of the resources?** Some informational resources may be evidence-based (e.g., they use content and communication strategies supported by scientific evidence), while others may not.
- **Are the resources sustainable?** This question addresses how likely it is that the informational resources will be available for continued use.

To answer these questions, information can be gathered by reviewing the informational resources themselves, as well as any documentation regarding the development of the resources, and by speaking with individuals who were instrumental in the development process.

Key questions related to evaluating the **processes** related to informational resources include

- **How are the informational resources disseminated?** For example, are resources being actively disseminated by email or postal mail to those on an organization's mailing list or by bringing informational resources to attendees at a community presentation? Can potential users find (and download) resources online?
- **Who is accessing the resources?** Because informational resources can vary dramatically according to their target audiences, objectives, and topics, it is important to understand who is obtaining and using them. This may necessitate gathering information about whether users fall into specific categories of interest (e.g., racial and ethnic minorities, certain job titles). It can also be helpful to collect data on how frequently resources are being accessed.

These questions can be addressed by collecting data on how resources were disseminated and to whom, as well as by interviewing individuals who play key roles in the dissemination process. Key questions related to evaluating the **outcomes** of informational resources include

- **Did the informational resources affect the targeted short-term outcomes?** Depending on the objectives of the informational resources, their use could influence knowledge, attitudes, and intentions to behave in a certain way (e.g., employers who read a fact sheet dispelling myths about employing people with mental health challenges may decide to change hiring policies).
- **What long-term outcomes do the informational resources affect?** Although the development and dissemination could affect key long-term outcomes, such as population levels of stigma and discrimination, it is difficult to trace such changes to a specific resource or resources. Therefore, we focus more on the short-term outcomes of informational resources.

These questions can be addressed by surveying users of informational resources.

Evaluation Design

Different strategies can be used to understand the structures, processes, and outcomes of informational resources. We discuss a number of design possibilities for evaluating these components. Not all of the design issues discussed below will apply to all types of informational resources. When considering an evaluation design, it is important to take into account the nature of the informational resource and its objective.

Structure

Two methods for better understanding informational resources are **document reviews** and **key informant interviews**. Reviewing informational resources can determine the topics they cover, including the general nature of the content, as well the breadth of content they cover. A review can also determine what audiences are being targeted, what strategies were used to tailor the resources to the target audiences, and whether the resources are evidence-based (i.e., supported by scientific research). Lastly, a review can help gauge the sustainability of resources (e.g., whether the materials are likely to become outdated quickly, funding sources).

A document review can be supplemented by key informant interviews—that is, interviews with people who played an instrumental role in the development of the informational resources.

Process

Methods for understanding the processes associated with the use of informational resources include **tracking who receives informational resources, conducting key informant interviews, and surveying users to whom resources are distributed.** Many resources are made available online, so website analysis tools (e.g., Google Analytics) can be used to understand how sites hosting informational resources are navigated by users. To learn more about the people obtaining the resources online, a registration process can ask users to provide information about themselves. If it seems likely that many users would wish to remain anonymous, linking to an anonymous survey about user characteristics presents an alternative to registration. Speaking with informants who played major roles in the development and dissemination of informational resources can also provide insight into how informational resources are made available.

Surveys administered to those who have obtained informational resources can serve several useful functions. As mentioned, surveys provide a way to collect data about user characteristics. Surveys can also include questions about satisfaction with the dissemination of informational resources, as well as with the informational resources themselves.

Outcome

Understanding the short-term outcomes potentially affected by informational resources involves considering what outcomes to measure, when to measure them, and to whom should measures be administered. Most short-term outcome measures are assessed in surveys to resource users.

What Outcome Measures to Assess

Outcome measures used to evaluate informational resources should be selected based on the objective and content of each individual resource. Table 1.1 presents outcome measures that could be assessed for informational resources with any one or more of the following objectives: reducing stigma and discrimination, preventing suicide, and improving student mental health. Although the table provides some suggested outcome measures for various mental health topics, the selection of outcome measures should be specific to the informational resource being evaluated. For example, if one informational resource is a video about how teachers can refer at-risk students to mental health resources, it would be logical to include measures of teacher attitudes toward and comfort with making referrals in a survey administered to teachers who viewed the video.

When to Assess Outcome Measures

As with the choice of what outcome measures to administer, the choice of when to administer outcome measures may also vary based on the nature of a particular informational resource. If an informational resource is intended to change attitudes, knowledge, and behavior, then administering a survey both prior to and after receiving the resource is ideal, as is administering the survey to a carefully selected control group (see Chapter Two, on evaluating TE activities, for a more detailed discussion of pre-post designs and control-group selection). However, including pre-post assessments may require too much time from resource users (e.g., including a baseline survey could deter users from actually downloading resources), or it may not be feasible to administer a baseline survey (e.g., if resources are mailed to participants; also, it

would be difficult to ensure that participants completed a survey both before and after reviewing the resource). In these cases, outcome measures may only be assessed after resources are obtained. Although outcome measures could be administered immediately after resources are obtained, it may be beneficial to administer them at a later point (either in addition to or instead of immediately after receiving resources). For example, if users access an informational resource consisting of a directory of evidence-based practices and programs, it may be more useful to wait several months to learn whether users implemented one of the programs in the directory, as opposed to learning immediately after directory access that users said that they were inclined to implement a program (given that a program would not have been immediately implemented).

Who Should Complete Outcome Measures

Outcome measures should be administered to those who have accessed informational resources. In addition, the measures should be appropriately tailored for the various audiences that might access the resources. For example, if informational resources about reducing mental health stigma and discrimination are developed for and disseminated to both teenagers and health professionals, outcome measures should be tailored for and administered to both groups.

Evaluation Data Collection Plan

Structure

Above, we suggested the use of document reviews and key informant interviews to address key evaluation questions about the structures developed as part of the creation of informational resources. See Chapter Two for a detailed discussion of data collection considerations associated with using document reviews and key informant interviews. We recommend reviewing materials for and interviewing key informants about the contents of the resources, the objectives of the resources, their formats (e.g., hard-copy materials, materials hosted online, video), target audiences, details on how resources were tailored for target audiences, resource consistency with the evidence base, and factors affecting sustainability (e.g., likelihood of becoming outdated, funding for resources, staffing for resources).

Process

We suggest tracking who receives informational resources, conducting key informant interviews, and either asking resource users to register or complete surveys as methods for evaluating informational resource–related processes. Tracking the dissemination of resources can be accomplished using online tools, such as Google Analytics, to track the frequency with which pages containing resources are viewed, how often resources are downloaded, and how often video resources are viewed. These data can then be supplemented with survey items that ask users to report the frequency with which they visit a site. Additional process-related variables, such as satisfaction and perceived utility of the resources, and the method of dissemination can be included in these user surveys. Further information about the dissemination of resources can be obtained by conducting key informant interviews, which have been discussed in greater detail in Chapter Two.

Information about user characteristics can be collected during a registration process or from user surveys. Characteristics most relevant to the specific informational resources being

evaluated should be assessed. These might include such characteristics as the race or ethnicity of the user or of the population with whom the user works (e.g., if the user is a mental health service provider working with specific populations), or the user's roles (e.g., teacher, student, service provider).

Outcome

We recommended administering user surveys to assess short-term outcomes. Considerations for collecting survey data are detailed in Chapter One. Outcome measures should include measures relevant to the objectives and nature of each informational resource (also see Chapter One for more information about outcome-measure selection).

Helpline Operations

This chapter provides an overview of how to approach evaluation design and implementation for helpline operations. We describe what kinds of activities are included in this category, review a logic model for helplines, and use the logic model framework to describe potential evaluation questions and designs for helpline operations. Finally, we discuss how to approach data collection for helplines.

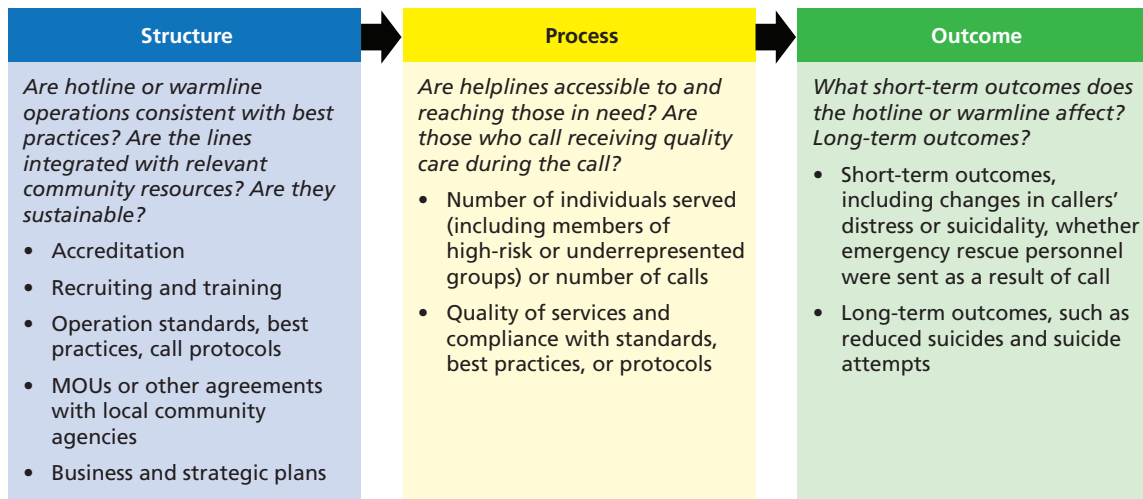
Description of Activity

Helplines are key activities used in suicide prevention and other PEI programs. *Hotlines* focus on crisis intervention, and *warmlines* focus on noncrises, often with the parallel goals of addressing other concerns while freeing up hotlines for emergencies. Suicide-prevention activities may focus on expanding the availability of helplines (i.e., hotlines and warmlines)—for example, by establishing new lines, expanding hours of existing lines, or making the lines accessible to more groups by adding operators who speak different languages. PEI activities may also focus on improving the quality of existing lines—for example, by meeting standards for accreditation or adding call monitoring or call recording to improve supervision and quality assurance.

Logic Model

Figure 4.1 presents a logic model to guide the evaluation of helplines. Structures can include capacities and resources that indicate consistency with best practices (such as accreditation, operation standards, and protocols for response to calls) and integration with the community, (such as MOUs or other agreements with local community agencies). Processes can include information on the reach of the lines to different groups of individuals (e.g., overall number of people served, members of high-risk or underrepresented groups served) and the quality of the care provided during the call. Although hotlines and warmlines often ultimately hope to affect key long-term outcomes, such as suicide rates, these effects can be difficult to observe. Thus, appropriate short-term outcomes might include immediate benefits to callers that can be rated by trained observers, such as the reduction in levels of distress, appropriate referral for services, or receipt of emergency response services if appropriate (see Jaycox et al., 2015, for detail on how these outcomes can be measured and a tool for monitoring the content of hotline calls).

Figure 4.1
Hotline and Warmline Operations Logic Model and Key Evaluation Questions



RAND RR1882-4.1

Key Evaluation Questions

Figure 4.1 includes a list of key evaluation questions that can guide the evaluation of the structure, process, and outcome of helplines. In this section, we provide further detail about these questions, and we describe the kinds of data that may be collected to answer the questions.

Key questions related to evaluating the **structure** of a hotline or warmline include

- **Are hotline and warmline operations consistent with best practices?** Are hotline and warmline operations following objective standards for best serving callers? It is tricky to answer these questions because there are currently no evidence-based standards to guide how hotlines and warmlines operate. However, various organizations have set forth standards for hotlines. Hotlines may be examined for accreditation or consistency with standards set forth by the American Association of Suicidology or other organizations (e.g., CONTACT USA, the Joint Commission, Alliance of Information and Referral Systems, Commission on the Accreditation of Rehabilitation Facilities). Hotline operations may also be assessed according to the National Suicide Prevention Lifeline's recommendations and criteria for joining its network (National Suicide Prevention Lifeline, 2007).
- **Are hotline and warmline operations integrated with relevant community resources?** The efficacy of a local hotline or warmline depends, to a certain extent, on its integration with the community. For example, does a hotline or warmline operation know about and make use of community resources for referrals and emergency rescues (and, conversely, are these community resources familiar with the hotline or warmline operation)?
- **Is the hotline or warmline operation sustainable?** This question is asked to understand the factors that enhance the likelihood of a hotline or warmline being able to continue into the future.

Different sources of information will be useful in answering these questions regarding structure. To evaluate consistency with best practices, it may be useful to evaluate accreditation

standards, as well as those established by the National Suicide Prevention Lifeline to identify minimum thresholds for performance, and then evaluate the hotline for meeting those performance standards. For hotlines that are accredited or members of the National Suicide Prevention Lifeline—or applying for accreditation or membership—this can be achieved through a review of application or reaccreditation materials. For lines that are not accredited, not members, and not planning on becoming accredited or joining the National Suicide Prevention Lifeline, crisis center staff can be interviewed regarding the helpline’s consistency with the standards. Samples of calls can be monitored live or recorded periodically to assess compliance with standards—e.g., the National Suicide Prevention Lifeline and the training protocol of the hotline or warmline.

To evaluate integration with community resources, it is important to examine MOUs or other cooperative agreements that the hotline or warmline has with local community agencies. It may be helpful to supplement review of official agreements with staff interviews, because there may be informal agreements in place (e.g., police departments may be reluctant to provide formal agreements, but there may be a strong understanding in place nonetheless).

To evaluate hotline and warmline sustainability, review of accreditation materials or interviews with staff may be used to estimate the cost of running the helpline and the funding mechanisms that currently support these operating costs. It is then possible to gauge whether these funding streams are sustainable. The assessment of sustainability can also take into account the strength of a helpline’s relationships with other community resources.

Key questions related to evaluating the **processes** involved in hotline and warmline operations include

- **Are helplines accessible to and reaching those in need?** Are people aware of and accessing hotlines and warmlines when they need help? Answering this question involves assessing not just the overall number of people the line reaches but also who the line reaches, including members of groups that may be particularly high risk or typically have less access to services.
- **Are those who call receiving quality care during the call?** While the first process question focuses on access to services, the second question focuses on the appropriateness of the services. For example, what kinds of techniques, such as active listening and collaborative problem-solving, are used to help callers?

Different data sources may be used for different process measures. Data on the population size and demographics of the hotline’s or warmline’s catchment area should be available. Data on relative risk in different population groups—e.g., the elderly—should also be available. To assess reach, data can be collected on the number of calls or callers, including the number of callers in different demographic groups and comparisons made with local demographics. Consistent and reliable demographic data collection from callers in emotional distress requires careful training of call-responder staff and monitoring for compliance. Wait times for inbound calls or number of calls rolled over to backup lines, as well as days and hours of operation and languages or translation services available, may help assess reach and access. Tracking the impact of new outreach efforts could also shed light on questions of access. To better understand accessibility from a consumer perspective, it may be possible to survey individuals in a hotline’s or warmline’s catchment area about their awareness of local helplines and why these individuals would or would not use a given resource. Collecting demographic information can

be helpful in understanding which groups are more aware of and more likely to use helpline services. However, this kind of surveying may be beyond the scope of what some lines can feasibly implement.

To assess quality of the calls, the source of data is the calls themselves, which can be evaluated using live call monitoring or analysis of recorded calls by trained reviewers according to standards for best practices.

Key questions related to evaluating the **outcomes** of a hotline or warmline include

- **Did the hotline or warmline affect the targeted short-term outcomes?** Did a hotline or warmline improve a caller's emotional state? Were appropriate referrals given? Did the caller receive immediate assistance, if necessary (i.e., emergency rescue for a imminently suicidal caller)? Answering these questions can help clarify whether callers are benefiting from their experience with the helpline.
- **What long-term outcomes are affected?** Helplines often have the long-term aim of reducing suicides or suicide attempts, but it can be difficult to measure changes in suicide rates in a community; further, it is not possible to determine whether overall changes in suicide rates are attributable to a particular helpline. Thus, we recommend that program evaluations largely focus on short-term outcomes, which are simpler to measure and easier to attribute to the helpline.

To address the short-term outcomes of the call, the content of the call itself is once again the source of data. Using a standardized protocol, in the form of a rating checklist, trained reviewers may conduct silent live monitoring of calls or review recorded calls and rate changes in callers' emotional states between the beginning and end of the call. Raters may also assess compliance with the helpline's training standards or National Suicide Prevention Lifeline standards, whether appropriate referrals were given, and whether emergency rescue personnel were sent as a result of the call (when appropriate).

Evaluation Design

In this section, we present some design considerations relevant to evaluating structures, processes, and outcomes of hotline and warmlines.

Structure

Answering key evaluation questions about hotline and warmline structures requires quantitative and qualitative methods, such as **document reviews** and **key informant interviews**.

Document Reviews

Materials, such as application or reaccreditation documents and training protocols, can be reviewed against accreditation and National Suicide Prevention Lifeline standards to evaluate whether the helpline meets these performance standards. Review of accreditation materials and strategic and business plans can provide insight into funding mechanisms and operating costs, which can be taken into account when evaluating the helpline's sustainability. MOUs and other cooperative agreements with local agencies may be reviewed to glean information about the degree of integration with community resources. Assessment of the resource direc-

tory used by call responders should also be conducted periodically to identify gaps and to ensure that resources included remain in business and have accurate contact information and hours of operation.

Key Informant Interviews

Interviews can be conducted with hotline and warmline staff who are knowledgeable about hotline and warmline operations. More specifically, it may be useful to conduct semistructured interviews, in which there is a defined set of interview questions but also the flexibility to deviate from the exact question wording and order. Interview questions might focus on the helpline's consistency with performance standards, relationships and agreements with other community resources and with other helplines, adequacy of referral resources, trends in staff recruiting and community outreach, and operating costs and funding mechanisms.

Process

To address questions about hotline and warmline processes, both qualitative and quantitative methods—such as collecting data on reach, surveys, and silent live monitoring or recording of calls—can be used.

Collecting Data on Reach

Demographic information on callers (e.g., gender, race and ethnicity, age, sexual orientation, geographical location) may be routinely collected during calls.

Surveys

Surveys may be used to query individuals in a hotline's or a warmline's catchment area regarding their awareness of the helpline and their willingness to use the resource. These data could be tabulated quantitatively, by demographic group. However, sampling a population of individuals is a complex task that would likely require expert consultation.

Silent Live Call Monitoring

In a process called *silent live monitoring*, both sides of calls can be monitored silently in real time by trained raters who compare the care provided with standards for best practices, based on a standardized protocol. These standards could include those set forth by accreditation agencies or the National Suicide Prevention Lifeline, as well as the relevant research literature on conducting suicide risk assessments and intervening according to existing theories related to active listening and collaborative problem-solving models (e.g., Gould et al., 2012; Mishara, Chagnon, and Daigle, 2007). Raters can also consider transcripts of chat conversations or texts, for helplines offering these services. An alternative method of recording calls and subsequently reviewing them can also be used and offers an advantage of having real calls to use in training and having the opportunity to do the assessment with more raters. A protocol, developed by RAND for a study of crisis-line call content and quality in California is available, without charge, in Jaycox et al. (2015).

Outcome

We focus on evaluation design for short-term, or proximal, outcomes because assessing long-term outcomes is beyond the scope of most program evaluations. Ramchand et al. (2016) provide a summary of research on short-term outcomes and the results of a RAND evaluation of ten hotlines in California.

Silent Monitoring

Silent monitoring may also be used to examine the short-term outcomes of calls. Call raters can assess a variety of different outcomes when they listen in on calls (or potentially review transcripts of chats or text messages or listen to recorded calls). Various considerations are described below.

Which Outcomes to Assess

The program should consider the goals of its hotline or warmline when deciding which outcomes are most important to assess. If the goal of a hotline is to reduce suicidality of callers, then this is an important outcome to assess. If the goal of a warmline is to connect individuals with other supportive services in their community, then it will be important to assess the referrals provided. In sum, it may be important to rate such factors as a caller's level of distress, suicidal ideation, suicidal intent, whether a safety plan was developed, whether appropriate referrals were given, and whether emergency rescue personnel were sent as a result of the call.

When to Assess Outcomes

In general, it is recommended that outcomes of interest be rated at both the beginning and the end of a call, so that changes over the course of a call can be observed. This applies to ratings related to the caller's emotional state and suicidality. Results of the call—such as referrals, safety plans, and rescues—can only be assessed once, at the end of the call.

Who Should Assess Outcomes

Ideally, outcomes are assessed by independent raters who are not affiliated with the hotline or warmline being evaluated. Staff of a given helpline may be biased in favor of viewing the calls as helpful, and it is best to obtain an objective evaluation of whether callers benefit from the call.

How to Ensure That Assessments Are Reliable

It is important to ensure that raters are consistent with each other and with the rating criteria. To ensure this, raters should receive training prior to rating calls and should rate practice calls (these could be real calls or mock calls that are role-played) until the raters are consistent with a standard. Once silent monitoring has begun, a subset of calls should be evaluated by two raters, so that their ratings can be compared to ensure that they are consistent with one another. This consistency between raters is referred to as *interrater reliability* and may be computed statistically.

Surveys

Surveys could also be used to follow up with hotline and warmline callers, to determine whether users made use of the referrals they received (e.g., sought help from a therapist). Surveys could be administered via phone, mail, or online. Permission must be obtained to follow up with callers at the time of the initial call, so follow-up surveys will be limited by callers' willingness to provide permission and contact information.

Evaluation Data Collection Plan

In this section, we present considerations for developing a data collection plan for the evaluation of helplines.

Structure

In the evaluation design section, we suggested using **document reviews** and **key informant interviews** for addressing key evaluation questions about the structure of helplines.

To conduct document reviews, materials must be collected, stored (online or in hard copy), and made available for conducting the evaluation. Then, a data abstraction tool must be designed to facilitate review of the materials. A common format for such a tool is an Excel spreadsheet, with rows for each material being reviewed and columns for the different kinds of information being obtained (i.e., abstracted) from the materials. There might be columns for various specific performance standards, cooperative agreements with local agencies, and funding mechanisms to inform questions regarding consistency with best practices, integration with community resources, and sustainability, respectively.

To conduct key informant interviews, appropriate informants must first be identified. If possible, it is best to identify multiple informants to get multiple perspectives on the relevant issues. An interview protocol must be developed that includes questions on all the topics of interest. It may also be helpful to develop a plan for taking notes on the interviews. Notes may be taken on the protocol itself, or it may be helpful to develop a separate response sheet that parallels the interview protocol. Sometimes, instead of taking detailed notes during interviews, evaluators elect to record the interviews and then either take detailed notes based on the recordings or have the recordings transcribed for later review. Interviewers may need to be trained on how to conduct interviews or how to take notes on them, and if multiple interviewers are used, care should be taken to ensure that their approaches to interviewing and documenting the interviews are consistent with one another. For more detail on qualitative methods, see Chapter One, particularly the resources listed in Table 1.5.

Process

To assess processes related to hotline and warmline operations, we suggested collecting data on reach, possibly administering surveys of community members, and conducting silent live monitoring or recording of calls.

As discussed, the reach of helplines can be assessed by comparing population size and characteristics (e.g., risk factors) of the hotline or warmline catchment area, with data on the number of calls and callers, including key information about the characteristics of callers (e.g., gender, race and ethnicity, age, sexual orientation, geographical location, and factors). Data collected during calls may be entered into helpline software that is commonly used to track calls (e.g., iCarol). This software can then be used to generate the total number of calls and callers in a given period, as well as break down calls by demographic group (quantitative data). If helpline software is not available, this kind of tracking can also be done via paper and pencil and tabulated in Excel spreadsheets. Careful training and supervision of staff are required to facilitate collection of the data and ensure that data are comprehensive and reliable across call responders.

These data can be compared with data on the need for hotline and warmline services among various groups (e.g., how many callers were from groups at high risk for suicide, how

many callers were from groups that tend to have less access to preventive services, and how many calls were handled in languages other than English?). The data can also be compared with the helpline's goals. For example, if a helpline had a goal of increasing services to a particular group through an outreach effort, the characteristic data of the call can be used to assess whether the number of calls from this group changed over time.

If surveys are administered to individuals in the community, the appropriate catchment area must be identified, a sampling strategy must be developed, a means of data collection must be chosen (e.g., phone, mail, online), and the survey questions must be designed. Again, this task would likely require expert consultation.

To conduct silent live monitoring of calls, a call sampling plan (e.g., one to three calls per responder per shift, depending on call volume, staffing, and other logistics), a protocol, and data collection tool must first be developed. (See Jaycox et al., 2015, for a copy of a protocol developed by RAND for a crisis-line quality assessment of ten hotlines based in California.) The tool should include basic information about the call rater (e.g., rater identification number), characteristics of the call (e.g., date, time, duration, type of call [crisis call, noncrisis call, prank call, and others]), and caller (e.g., gender, other demographic characteristics ascertained). Names of callers and call responders should not be entered on the data collection tool. Further, the tool should include items about the standards of interest (e.g., call responder behavior, such as active listening, collaborative problem-solving).

Call raters should be trained, and their consistency with objective standards and with one another should be established prior to implementation of call monitoring, as described in the design section above. They should also be trained on procedures to follow if they hear a call that suggests imminent risk for suicide and the call responder is failing to follow the crisis line's protocol.

The purpose and plan for call monitoring should be explained to all call responders who will be monitored as an organizational effort to improve the quality of services and identify areas for additional training for new and existing staff. Responders should understand that only some calls will be monitored or recorded and that they and their callers will not be aware of the raters on the line. They may be concerned that the callers will object to having calls monitored over concerns of privacy, but prior studies using call recording and call monitoring have not experienced significant rates of caller refusal or disconnected calls. However, in the rare instance in which a caller objects, the staff should be trained to know that the call rater will drop off the call and that they can tell the caller that the call will not be monitored. During periods of silent monitoring (or call recording), call responders should be trained to provide a brief, one-sentence statement to callers explaining that calls may be monitored or recorded for quality purposes. Responders should be trained to move on quickly to the caller's reason for calling but have ready answers to questions that a few callers might have about the monitoring or recording. Some hotlines and warmlines have recorded greeting messages when a call comes in and can add a statement about call monitoring or recording as an alternative to asking the call responders to read the message. Hotlines and warmlines considering call monitoring or call recording should consult the regulations of the Federal Communications Commission (FCC) and their states' communications regulatory agencies for rules regarding notification of call monitoring or recording.

Outcome

Although it is possible to use surveys to assess outcomes, we suggest using silent live call monitoring or recording as the primary means for collecting data on outcomes. The call outcome assessment protocol should include all the basic information described in the process section above, as well as items about outcomes of interest (e.g., caller satisfaction, a caller's emotional state and suicidality at the beginning and end of the call, whether a safety plan was developed, whether appropriate referrals were given, and whether an emergency rescue was appropriately initiated). An example of a live call monitoring protocol developed by RAND for an assessment of crisis line call content and quality is available in Jaycox et al. (2015).

References

- Gould, M. S., J. L. Munfakh, M. Kleinman, and A. M. Lake, "National Suicide Prevention Lifeline: Enhancing Mental Health Care for Suicidal Individuals and Other People in Crisis," *Suicide and Life-Threatening Behavior*, Vol. 42, No. 1, 2012, pp. 22–35
- Jaycox, Lisa H., Rajeev Ramchand, Pat Ebener, Dionne Barnes-Proby, and Marylou Gilbert, *RAND's Silent Monitoring Protocol for Assessing Suicide Crisis Line Call Content and Quality*, Santa Monica, Calif.: RAND Corporation, 2015. As of April 26, 2017:
<http://www.rand.org/pubs/tools/TL150.html>
- Mishara, B. L., F. Chagnon, and M. Daigle, "Which Helper Behaviors and Intervention Styles Are Related to Better Short-Term Outcomes in Telephone Crisis Intervention? Results from a Silent Monitoring Study of Calls to the U.S. 1-800-SUICIDE Network," *Suicide and Life-Threatening Behavior*, Vol. 37, No. 3, 2007, pp. 308–332.
- National Suicide Prevention Lifeline, *Suicide Risk Assessment Standards*, New York, April 17, 2007. As of April 26, 2017:
<https://suicidepreventionlifeline.org/wp-content/uploads/2016/08/Suicide-Risk-Assessment-Standards-1.pdf>
- Ramchand, Rajeev, Lisa H. Jaycox, Pat Ebener, Marylou Gibert, Dionne Barnes-Proby, and Prodyumna Goutam, "Characteristics and Proximal Outcomes of Calls Made to Suicide Crisis Hotlines in California," *Crisis*, Vol. 38, No. 1, 2016, pp. 26–35.

Networks and Collaborations

This chapter provides an overview of how to approach evaluation design and implementation for networks and collaborations. We begin by describing what kinds of activities are included in this category. We then review what a logic model for networking and collaboration activities might look like, and, based on this logic model, we go on to describe some key evaluation questions for these interventions. We then review what evaluation design might look like for networks and collaborations and how to approach data collection for this type of activity.

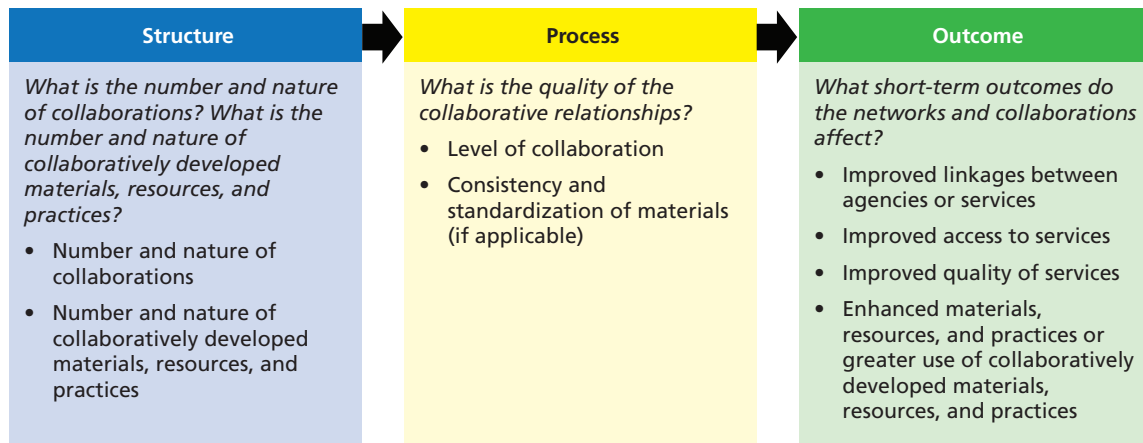
Description of Activity

Networks and collaborations are partnerships between organizations to work toward cross-system goals. These collaborative structures may serve various purposes. In many cases, collaborative networks can facilitate the sharing and dissemination of ideas, knowledge, resources, or standardized practices. Similarly, collaborations can enable agencies to coordinate the services they provide. Further, collaborations enable programs to build referral networks to increase individuals' access to needed services. For example, a student mental health policy workgroup in California was established to develop statewide policies that would affect K–12 programs and to promote linkages between schools and other organizations to better serve students (California Department of Education, 2017). Networks and collaborations are different from the other activities described in this handbook because, rather than occurring within a single organization, they occur across organizations. Therefore, this activity has its own unique challenges with respect to program evaluation.

Logic Model

Figure 5.1 contains a logic model to guide the evaluation of networks and collaborations. Structures include the number and nature of collaborations, as well as the number and nature of collaboratively developed materials, resources, and practices. Processes reflect the depth of the collaborative relationship and may include the consistency of collaboratively developed materials. Outcomes are the changes resulting from networking and collaboration activities. We provide some examples, but the outcomes of interest will depend on the purpose of the networks and collaborations and must be tailored to the goals of a program.

Figure 5.1
Networks and Collaborations Logic Model and Key Evaluation Questions



RAND RR1882-5.1

Key Evaluation Questions

It is important to consider the questions you hope the evaluation will answer, because they will guide the evaluation design. In this section, we provide some examples of key evaluation questions that programs may wish to answer regarding structure, process, and outcome for networks and collaborations (also see Figure 5.1). Note that the evaluation questions will vary quite a bit from program to program, depending on the goals of the networks and collaborations, particularly with respect to outcomes, so these questions should be considered as examples that are not meant to be comprehensive.

Key questions related to evaluating the **structure** of networks and collaborations include

- **What is the number and nature of the collaborations?** This question seeks to understand how many and what type of collaborative relationships were developed or enhanced. For example, did your organization increase the diversity of programs it collaborated with, or did it increase the number of connections to other similar programs? Were existing relationships strengthened, or were new collaborations established?
- **What is the number and nature of collaboratively developed materials, resources, and practices?** Some programs may link to similar programs for the purpose of sharing materials, resources, and practices, and this question aims to describe the number and nature of these shared items. You will want to document and categorize these shared materials (e.g., whether they are training resources or best-practices guidelines).

Questions regarding collaborative structures can be answered by reviewing relevant documents, conducting key informant interviews, and administering surveys, as described below.

One key **process** question related to evaluating networks and collaborations:

- **What is the quality of the collaborative relationships?** This question seeks to evaluate the level of collaboration between the program of interest and its partners, including the extent of connections, how closely organizations are collaborating within networks and communities, and the sustainability of the connections. Where applicable, it may also be

important to evaluate the consistency and standardization of materials, resources, and policies generated by the collaborative partnerships. If a goal of the collaborations is to facilitate referrals, it is also appropriate to evaluate the perceived ease and flexibility of the referral processes among members of the collaboration.

Understanding the processes associated with networks and collaborations may involve document review, key informant interviews, and surveys, as described below.

A key question related to evaluating the **outcomes** of networks and collaborations:

- **Did the network or collaborations affect the targeted short-term outcomes?** Short-term outcomes are the changes that occur as a result of the collaborative relationships. For example, an evaluation might ask whether close collaborative relationships resulted in increased coordination of services and increased access to services. For example, are there increased referral rates? Increased use of services? Also, have the collaborative partnerships resulted in enhanced materials, resources, and practices or greater use of collaboratively developed materials, resources, and practices?

We highlight short-term outcomes because longer-term outcomes, such as improved mental health, are difficult to specifically link to enhanced networks and collaborations.

Understanding the short-term outcomes of networks and collaborations primarily involves conducting surveys, interviews, and focus groups with key informants; document review of materials and resources developed through the collaborative partnerships; and quantitative analysis of referral rates and service utilization rates when increased referrals and service utilization are outcomes of interest.

Evaluation Design

A number of design considerations related to evaluating networks and collaborations are presented below. However, specific methods of evaluating networks and collaborative relationships should be selected based on the nature and goals of the collaborations being evaluated. It is particularly important to adapt the design to the specific outcomes of interest.

Structure

Key approaches for evaluating collaborative structures include **document review**, **key informant interviews**, and **surveys**. Reviewing MOUs and other collaborative agreements, meeting rosters and attendance logs, and lists of network members can help quantify the number and types of collaborations in place. Reviewing collaboratively developed materials and resources can help determine the number and nature of these materials, including their objectives. Key informant interviews, focus groups, or surveys can help determine the number of collaborative connections established, with whom the connections were made, and the primary functions of the collaboration.

Process

Understanding the processes associated with networks and collaborations may also involve **document review**, **key informant interviews**, and **surveys**. A **document review** of collabora-

tive agreements and MOUs can provide information about the nature and extent of organizational connections. From these agreements, you can document the scope of collaborations in terms of reach (e.g., overlap or division of community coverage), funding (e.g., agreements to share or pool funds), or frequency of contact.

Key informant interviews can be used to determine the extent to which collaborative partners are making connections, the manner in which these collaborative relationships are being established or maintained, and the perceived quality of the connections being made. The quality of a collaboration can be evaluated on a number of dimensions, including how frequently contact is made between collaborative partners, the manner in which collaborative partners interact, the level of collaboration, and the reciprocity of collaborative partnership (e.g., who initiates contact). Focus groups may be used in lieu of individual key informant interviews. Surveys may also be used to assess these same dimensions, instead of or in addition to key informant interviews. Surveys have the advantage of providing more-standardized information, while key informant interviews have the advantage of providing richer, more-detailed information.

Outcome

Evaluating the short-term outcomes of networks and collaborations may involve conducting **surveys; key informant interviews and focus groups; reviews** of materials and resources developed through the collaborative partnerships; and **analyses of referral and service utilization rates**, when increased referrals and service utilization are outcomes of interest. Again, the appropriate design will depend on the goals of the networks and collaborations.

Key informant interviews or focus groups can measure perceptions of improved availability and accessibility, as well as the quality of services, the degree of coordination of services, and any reduction of duplication. In designing such interviews, it is important to assess the extent to which respondents attribute changes to the networking and collaboration activities. Surveys may once again be used instead of or in addition to key informant interviews, depending on whether it is desirable to have more-standardized information (surveys); richer, more-detailed information (interviews); or a combination of the two.

Reviewing documents can help assess whether the collaboration led to enhanced materials, resources, and practices, while interviews can be used to assess whether there is greater use of collaboratively developed materials, resources, and practices.

Finally, if collaborations aimed to increase these outcomes, it may be important to collect and analyze data on referral and service utilization rates.

When to assess outcome measures: Collaborative relationships take time to fully develop. As a result, it may be necessary to evaluate outcomes six months or longer after the collaborative relationships are established.

Who should complete outcome measures: Individuals with knowledge of the networking and collaborative activities that their organization is engaged in should participate in the interviews and surveys.

Evaluation Data Collection Plan

In this section, we present data collection considerations for evaluating the structures, processes, and outcomes associated with networks and collaborations.

Structure

Data collection techniques for evaluating collaborative structures include document review, key informant interviews, and surveys. To conduct a thorough document review, the program should collect and archive relevant documents on an ongoing basis, including collaborative agreements with other agencies; lists of member of networks and collaborations; agendas, meeting notes, and rosters from collaborative meetings; and collaboratively developed materials. The evaluator should create a data abstraction tool for systematically collecting information on these materials. A typical format would be an Excel spreadsheet, in which there is a row for each document reviewed and a column for each piece of information that is abstracted from the document. For examination of the number and nature of collaborations, columns might include the type of agreement (e.g., MOU, contract, informal), members of the collaboration, the purpose of the collaboration, and details on meetings. For collaboratively developed materials, columns might reflect the purpose or objectives of the materials, the storage (e.g., online, hard copy), and the intended audience for the materials.

Key informant interviews, focus groups, and surveys may also be used to examine the number and nature of collaborations. Key staff at the collaborative organizations may be queried about who they are collaborating with and the purpose of the collaborations. A first step is to identify appropriate staff to complete interviews or surveys, keeping in mind that it is helpful to have multiple informants to obtain multiple perspectives and make sure that important information is not omitted. An interview protocol or survey must then be developed. A survey can include both open-ended and close-ended questions. Close-ended questions provide data that are easier to interpret. In contrast, open-ended questions provide richer information but require that an evaluator interpret the data, often through the development of a coding scheme. See Chapter One for more information on using surveys for data collection.

If interviews are used, it may be helpful to develop a plan for taking notes on them. Alternatively, evaluators may choose to record the interviews and then either take notes on the recordings or have the recordings transcribed for later review. Interviewers may need to be trained on how to conduct interviews or how to take notes on them. If multiple interviewers are used, care should be taken to ensure that their approaches to interviewing and documenting the interviews are consistent with one another.

Process

Similar to structures, we suggest using document review, key informant interviews, or surveys to evaluate processes. The data collection methodology for evaluating processes would be the same as that described in the structure section in this chapter, but the type of information collected would differ (e.g., interview items relevant to processes would focus on the quality of the collaboration), as described in the evaluation design section. A single interview or survey may be developed and implemented to assess structures, processes, and outcomes to decrease the time burden on respondents.

Outcome

As with structures and processes, document review, interviews, and surveys may be used to evaluate outcomes of networks and collaborations. See the evaluation design section in this chapter for details on the factors that should be measured (e.g., perceptions of improved availability, accessibility, quality of services), and see the structure section for some considerations related to data collection.

We also recommend analysis of referral and service utilization rates when these factors are outcomes of interest. Examining these outcomes involves setting up a tracking system, if one is not currently in place. Typically, there would be a database in which these factors are tracked at the level of the client using the services, and the database would indicate what specific kinds of referrals or services were received.

References

California Department of Education, "Student Mental Health Policy Workgroup," web page, last reviewed May 1, 2017. As of May 3, 2017:
<http://www.cde.ca.gov/ls/cg/mh/smhpworkgroup.asp>

Social Marketing Campaigns

This chapter provides an overview of how to approach evaluation design and implementation for social marketing campaigns [SMCs]). We begin by describing what mean by *social marketing campaigns*. We then review a logic model for social marketing activities and use the logic model framework to describe key evaluation questions and evaluation design approaches for these activities. Finally, we discuss how to approach data collection for evaluation of the campaigns.

Description of Activity

SMCs are activities that use marketing techniques, such as websites and radio or television advertisements, to promote messages related to prevention and early intervention (e.g., to reduce the stigma and discrimination associated with mental illness and prevent suicide). SMCs may target any number of audiences, including youth, people with mental health challenges, and people who work in journalism and entertainment media.

Logic Model

Figure 6.1 presents a logic model to guide the evaluation of SMCs. SMC-related structures include the actual components of the campaigns, such as the messages created and disseminated across various channels (e.g., television, radio, Internet). Processes include the dissemination of SMC messages to target audiences. Outcomes are both the short- and long-term effects of SMCs, such as increased knowledge and awareness about mental health issues and reductions in the number of suicide attempts.

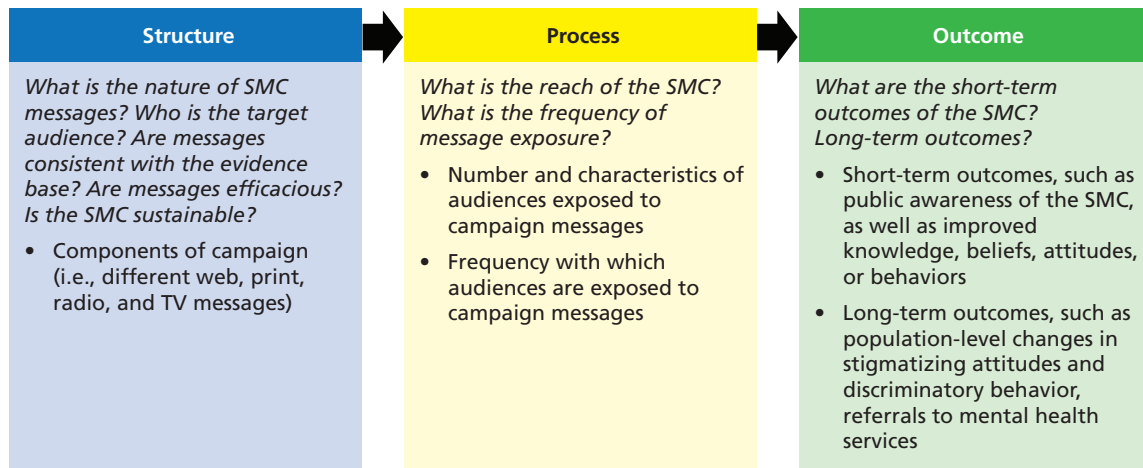
Key Evaluation Questions

Figure 6.1 also includes a list of key questions for evaluating the structure, process, and outcome related to SMCs. We summarize each key evaluation question below and briefly discuss what data can be used to answer the questions. See Chapter One for more-detailed information about evaluation design and data collection.

Key questions related to evaluating the **structure** of SMCs include

- **What is the nature of the SMC messages?** This question addresses both the content and format of SMC messages. Messages can vary dramatically on both of these dimensions

Figure 6.1
Social Marketing Campaign or Media Production Logic Model and Key Evaluation Questions



RAND RR1882-6.1

and can range from suicide-prevention advertisements aired on television to website messaging about reducing mental health stigma.

- **Who is the SMC's target audience?** Addressing this question requires knowing for whom the messages were developed and what strategies were used to tailor the messages for the target audience. For example, a stigma-prevention campaign targeting young people may include messages delivered in a colloquial language commonly used by that age group.
- **Are the messages consistent with the evidence base?** This question determines whether messages are designed and delivered in ways that are consistent with the evidence base (i.e., research literature).
- **Are the messages efficacious?** Exploring this question entails determining whether the messages are capable of bringing about the desired change in the intended audience. For example, if a message is designed to change attitudes about people with mental health challenges, then it is important to know whether it was designed in a way that is likely to result in short-term attitude changes.
- **Is the SMC sustainable?** This question addresses how likely it is that the SMC and associated messages will be available for use over time.

Answering these questions will require gathering information about SMC message content and speaking with individuals responsible for developing the messages. It is also helpful to consider the research evidence base for the message. If there is no prior research, experimental tests of message efficacy (described further below) may be useful, if they are feasible.

Key questions related to evaluating the **processes** involved in SMCs include

- **What is the reach of the SMC?** An SMC cannot influence people it does not reach, so it is essential to understand who is exposed to the campaign.
- **What is the frequency of message exposure?** In addition to understanding *who* is exposed to SMC messages (i.e., reach), it is important to understand *how frequently* individuals are exposed. A message that is heard once is not likely to have the same impact as one heard multiple times (Noar, 2006).

Data on the reach and frequency of SMC messages can be collected by tracking the rollout of SMCs and from advertising measures (e.g., impressions—that is, the number of people or homes where a campaign advertisement was shown), as well as from surveys of those potentially exposed to the message.

Key questions related to evaluating the **outcomes** of SMCs include

- **Did the SMCs affect the targeted short-term outcomes?** Answering this question enables an understanding of the *immediate* effects of SMCs on knowledge, attitudes, and intentions to behave in a certain way.
- **What long-term outcomes do SMCs have?** Long-term outcomes can result from SMCs, such as population-level changes in knowledge about, attitudes toward, and behavior toward people with mental health challenges, but these outcomes are challenging for individual programs to assess (see Collins et al., 2012). As a result, we focus here on short-term outcomes.

Data gathered through surveys of those exposed to SMCs are the key source for understanding the short-term outcomes of SMC messages.

Evaluation Design

Various design issues must be considered when evaluating SMCs, and strategies may vary depending on the nature of the campaign, the types of messages used, and the ways in which those messages are disseminated. Next, we present several design issues to consider.

Structure

Evaluating the structure of an SMC can best be conducted through **qualitative reviews of message content** (and subsequent comparison to the evidence base related to that topic), **key informant interviews**, and **experimental tests of message efficacy**. Campaign messages can be reviewed for their nature and content and for consistency with the evidence base related to the message content. For example, messages for suicide-prevention campaigns can be reviewed and compared with established criteria for safely reporting and depicting suicide (i.e., Reporting on Suicide, available at www.reportingonsuicide.org). Key informant interviews can be conducted to get more information about message content, the target audience for the campaign, and factors affecting the sustainability of the SMC.

Experimental tests of message efficacy can be used to determine whether it is even possible for a message to change outcomes among the target audience, regardless of the level of SMC exposure. For these experiments, small samples from target populations can be randomized to exposure to one of the SMC messages or to a comparison message that does not directly address the issues targeted by the SMC. Outcome variables of interest (see the evaluation design section on outcomes) can be assessed immediately after exposure to the message in both groups. Findings on the outcome variables of interest can be compared across the two groups to determine the immediate impact of the messages when audiences are directed to view them. More-desirable findings on the group exposed to the message relative to the control group indicate that the message can effectively change the outcomes of interest.

Process

Understanding the processes of implementing SMCs may involve **tracking information about SMC campaign dissemination, surveys of target audiences, and key informant interviews**. Understanding the reach of an SMC can be determined through various types of tracking data, including advertising measures (e.g., impressions for aired SMC messages) and audience metrics from such sources as Nielsen and Arbitron that indicate key characteristics of audience members. These data indicate who was exposed to SMC messages and, on average, how often. The audience metrics also indicate whether messages effectively reach the targeted audience. Data on other channels of dissemination (e.g., using web analytic tools to track online campaign elements, asking mental health advocacy groups to track the frequency with which they refer service users to SMC campaign elements) can also be tracked. These tracking data can be supplemented with surveys of target audience members, in which respondents are asked to describe their exposure to the SMC and its messages and with key informant interviews with SMC staff responsible for developing and overseeing the dissemination plan for the campaign.

Outcome

Measuring the short-term outcomes of SMCs requires consideration of what outcomes to measure, when, and from whom. As with many other types of PEI activities, surveys of the target audience are a key data source for measuring outcomes.

What Outcome Measures to Assess

The selection of outcome measures is highly dependent on the nature of the SMC. For example, a suicide-prevention SMC may feature messages that promote intervening with people in crisis. In this case, attitudes and beliefs about intervening, as well as perceived self-efficacy for doing so, are critical outcomes to assess. For more on possible outcomes to assess depending on the objective of the SMC, see Table 1.1.

When to Assess Outcome Measures

Because SMC messages are likely to reach members of the target audience at different times, outcome-measure assessment should occur at multiple points in time (i.e., a *time series* design; for more information, see Hornik, 2002; Kumkale and Albarracín, 2004; Lau and Russell, 1980; Noar, 2006). Measuring outcomes at multiple points can be helpful in distinguishing effects of the campaign from effects of outside events, and measurements extending after implementation is completed enables the detection of delayed effects. These measurements also allow for a comparison between the effects of the campaign and the effects of media trends in topics related to the campaign but not directly influenced by the campaign. For example, if an SMC is ongoing, it is important to note when an event involving a person with suspected mental illness receives major media coverage and to try to separate the effects of the campaign from the effects of the event unrelated to the campaign.

Who Should Complete Outcome Measures

Surveys of outcome measures should be administered to members of the target audience. This can be a challenging task, however, for large-scale SMCs. For example, a statewide campaign that involves radio and television ads discouraging teenagers from stigmatizing people with mental illness will require a statewide survey of teenagers. Administering a survey to a smaller sample could decrease the likelihood of detecting changes in outcome measures. Also, choos-

ing a smaller sample could increase the risk of biasing findings by focusing on a group that may not be representative of the entire target audience. For example, in the case of the statewide campaign, sampling only teenagers from Los Angeles presumes that their responses to outcome measures are likely to be representative of all teenagers in the state of California, which may not be the case.

Evaluation Data Collection Plan

In this section, we present considerations for collecting data necessary to answer the key evaluation questions surrounding SMCs.

Structure

Evaluating the structure of an SMC can require **qualitative reviews of messages**, **key informant interviews**, and **experimental tests of message efficacy**. Detailed discussions of qualitative reviews and key informant interviews were discussed in Chapter One. For SMCs, we recommend reviewing messages and asking key informants questions about the topic and content of a message, its mode of dissemination (e.g., television advertisement, poster), the message's intended target audience, and factors related to sustainability (e.g., funding related to the campaign). Collecting data for experiments testing message efficacy involves many of the same concerns as collecting survey data (discussed in Chapter One). For example, such factors as whether questionnaires should be administered at multiple times, who should be recruited to participate in the experiment, and what measures should be included on the questionnaire must all be considered. Outcome measures used in these experiments should be similar to those used in assessing the outcomes of the SMC (see the "What Outcome Measures to Assess" section earlier in this chapter).

Process

Testing SMC processes may involve tracking information about campaign dissemination, surveying target audiences about message exposure, and conducting key informant interviews. Tracking campaign dissemination information requires the collection of data for all means of dissemination. For example, if hard-copy materials are distributed, one source of data could be a log of which materials were sent, how many were sent, and to whom. Web analytic metrics, such as the number of times each page is viewed and the number of times a file is downloaded, and media metrics, such as the number of impressions associated with a television public service announcement, are other examples of dissemination data that could be collected for aspects of the campaigns disseminated online or on television, respectively. Other data sources could be identified for each channel by which a campaign is disseminated (e.g., social media, in-person presentations). Data collection considerations for conducting surveys and key informant interviews were presented in detail in Chapter One. For large-scale SMC campaigns, partnering with an organization or vendor experienced in recruiting representative samples of members of the target audiences and managing and analyzing large data sets may warrant consideration.

Outcome

Collecting data on outcomes of SMCs requires surveys of the target audience. Again, large-scale campaigns may require partnering with other organizations to successfully execute the collection of required data.

References

- Collins, Rebecca L., Eunice C. Wong, Jennifer L. Cerully, Dana Schultz, and Nicole K. Eberhart, *Interventions to Reduce Mental Health Stigma and Discrimination: A Literature Review to Guide Evaluation of California's Mental Health Prevention and Early Intervention Initiative*, Santa Monica, Calif.: RAND Corporation, TR-1318-CMHSA, 2012. As of May 3, 2016: http://www.rand.org/pubs/technical_reports/TR1318.html
- Hornik, Robert C., "Exposure: Theory and Evidence About All the Ways It Matters," *Social Marketing Quarterly*, Vol. 8, No. 3, 2002, pp. 31–37.
- Kumkale, G. Tarcan, and Dolores Albarracín, "The Sleeper Effect in Persuasion: A Meta-Analytic Review," *Psychological Bulletin*, Vol. 130, No. 1, 2004, pp. 143–172.
- Lau, Richard R., and Dan Russell, "Attributions in the Sports Pages," *Journal of Personality and Social Psychology*, Vol. 39, No. 1, 1980, pp. 29–38.
- Noar, Seth M., "A 10-Year Retrospective of Research in Health Mass Media Campaigns: Where Do We Go from Here?" *Journal of Health Communication*, Vol. 11, No. 1, 2006, pp. 21–42.

Screening and Referral Services

This chapter focuses on evaluation design and implementation approaches for mental health screening and referral services. We describe what kinds of activities are included in this category, review a logic model for screening and referral activities, describe some key evaluation questions, and review how you might go about answering them.

Description of Activity

The primary activities of screening and referral programs are the identification and assessment of individuals at risk of, or exhibiting early signs of, mental health challenges and the connection to appropriate resources or treatment. Examples of screening and referral programs include mobile screening programs (such as health fair screenings); programs in which peer advocates provide to their communities information about available mental health resources; and integrated primary care, mental health, and alcohol- or drug-screening programs.

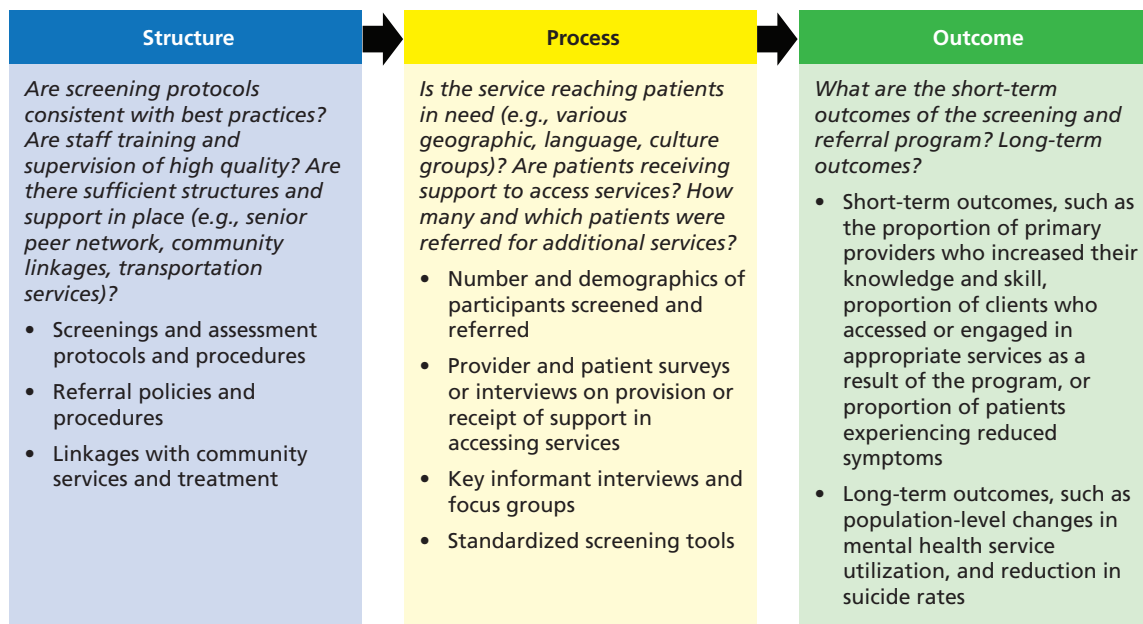
Logic Model

Next, we depict some common features of the structures, processes, and outcomes of relevance to screening and referral programs. Figure 7.1 presents a logic model with some common examples of relevant structures, processes, and outcomes. Components of the screening and referral services structure include screening and assessment protocols and organization referral policies. Participant records, provider and patient surveys, and key informant interviews can be used to understand the activity processes. Outcomes include the proportion of primary providers who increased their knowledge and the proportion of clients who accessed appropriate services.

Key Evaluation Questions

In this section, we give several examples of questions related to a program's structure, process, and outcome. Your evaluation questions will depend on the type of information you are interested in and your organization's capacity to collect and analyze it. No single evaluation will be able to answer all of the questions suggested here. You will need to identify those that are most relevant to your programmatic needs. For screening and referral programs with limited evalu-

Figure 7.1
Screening and Referral Logic Model



RAND RR1882-7.1

ation resources, evaluations are often centered on specific process questions (e.g., how many people were referred) or outcome questions (e.g., the proportion of patients accessing appropriate services as a result of referral) that address the totality of individuals served by the program.

Structure

Some key questions related to the **structure** of a screening and referral program include

- **Are screening protocols consistent with best practices?** This question asks whether clients are being screened using the best-available objective screening protocols. It is important to note that best-practice protocols and screening tools may vary by the population being screened (i.e., may differ by culture, age, language, and so on). You should seek to evaluate your program relative to the best practices for your target population.
- **Are staff training and supervision of high quality?** It is important to assess not only the quality of the services that staff members provide but also the quality of the training and support they themselves receive. Assessing the quality and consistency of training and supervision across staff can help you pinpoint where and why there may be differences in the screening and referral procedures that staff members follow or the care they provide.
- **Are there sufficient structures and support in place?** This is critical contextual information for a screening and referral program. After all, the ability of individuals to access the programs they are referred to can depend on the availability of supportive structures in the community, such as senior peer networks, linkages to community resources (e.g., child care), and transportation services.

Process

Some questions related to the **process** of a screening and referral program include

- **Is the service reaching patients in need, and are they receiving support to access services?** First and foremost, you may be interested in who is being screened and referred through your program. You should consider assessing the age, gender, and race and ethnicity of program participants, as well as other factors that may affect the accessibility of services (such as primary languages spoken).
- **How closely do participants match the intended audience?** This question asks whether the population you identified as being the target of your program is indeed the one using the screening and referral services.
- **How many and which patients were referred for additional services?** Once clients access your program, what are their trajectories? You can examine overall percentages of the number of clients who are referred relative to those who are not. It may be even more useful to compare the demographics and traits of individuals who are referred versus those who are not, a process that will help illuminate where there may be under- or over-representation. These findings will also help you understand and contextualize your subsequent outcome analyses.

Outcome

Short-term outcomes can include changes in the knowledge, behaviors, well-being, and help-seeking actions of participants engaged in the screening and referral process. For screening and referral programs, some evaluation questions on **outcomes** include

- **What are the short-term outcomes of the screening and referral program?** One measure of short-term outcomes is the proportion of screeners who increased their knowledge about early screening and referral procedures. If one of the goals of your program is to increase the ability of staff to screen for signs of mental illness and refer individuals for treatment to mental health care providers, you will want to collect data on the number of providers whose knowledge improved as a result of your program. Another measure is the proportion of patients who access and engage in appropriate services as a result of the screening and referral program. An important goal of screening and referral programs is to determine whether clients, once they have been screened and referred, are more likely to seek treatment and services. A third measure is the proportion of participants who experienced reduced symptoms or improved functioning? Although the main goal of a screening and referral program may be to engage clients in appropriate services, you may want to consider tracking whether clients see improvements in symptoms and functioning as a result of accessing services. You will want to compare your outcome data with your benchmark goals to determine where your program is succeeding or may need improvements.
- **What are the long-term outcomes of the program?** Long-term outcomes may include population-level changes in mental health service utilization and reduction in suicide rates. However, evaluation of such outcomes is beyond the scope of the current handbook.

Evaluation Design

Structure your evaluation design to answer the specific questions you identified. When selecting a design, consider what you want to measure, how you will measure it, who will be measured, and when the measurements will occur. The types of evaluation designs described below are possible suggestions for screening and referral programs; which ones you select will depend on your program's specific goals and the evaluation resources you have available.

Structure

For program structures, often the most useful evaluation design is a **descriptive** one. You want to be able to describe what resources exist within your program to understand how they influence your process and outcomes. To answer questions about your program's structure (such as, "Are screening protocols consistent with best practices?"), you may consider a thorough **document review** of your program's screening, assessment, or referral protocols and procedures. **Key informant interviews** with individuals responsible for developing or selecting organization policies and procedures will help you understand how and when these tools are used. Likewise, key informant interviews and surveys may be useful in gathering data about how staff make referral decisions and where linkages with community services and treatment exist. A **literature review** will help you determine which best practices apply to your particular target population. A **qualitative analysis** of program protocols and procedure documents will enable you to compare them with established best practices.

Process

As with evaluation designs of structure, process designs are often **descriptive**. You may want to include a **document review** of client records, including client demographics. This will help you answer important questions about the people you are reaching (i.e., the demographics and characteristics of the clients that are being screened and referred). To understand the interactions that take place during the screening and referral process, you will want to consider using provider and patient **surveys** and **interviews** to gather data from both perspectives. **Key informant interviews** or separate **focus groups** with providers and patients may help you develop a framework for understanding the processes at play and the key issues that are relevant to each group.

Outcome

Outcome evaluation designs often consist of **experimental**, **quasi-experimental**, and **pre-post survey** (before and after) methods. It can be difficult to say with confidence that a particular program caused an outcome, because there are many factors at play in any given environment that can contribute to or detract from program effects. The only way to show true causality is via experimental methods. However, for a screening and referral program, it is not usually feasible to conduct a randomized control trial (for more on experimental methods, see Chapter One); instead, you may consider a **quasi-experimental** design. Using this method, individuals who were screened and referred through your program might be compared with demographically similar individuals who were *not* (i.e., individuals who did not receive the intervention, such as clients at a neighboring organization that does not have a screening and referral program). With this approach, you might be able to show a difference in the propor-

tion of clients at your organization who experienced reduced symptoms when compared with clients who did not have access to the program.

Pre-post surveys can also be conducted before clients are screened and then again after they have been screened and referred, to determine any changes in individual outcomes. Similarly, you might use pre-post surveys with providers to assess changes in their knowledge or skill. Pre-post surveys cannot show causality but nevertheless can provide compelling information about changes over time.

Evaluation Data Collection Plan

Structure

Particularly for screening and referral programs, which are likely to be more invested in process or outcome evaluations, **key informant interviews** can be a relatively inexpensive and rich source of data on program structure. For example, a critical structure question for a screening and referral program may investigate whether there are sufficient supportive structures in the community where individuals can access services. To collect this information, you might interview key program staff who work with community partners and facilitate program connections. Likewise, if you want to understand the role of staff supervision in ensuring that clients are screened appropriately, you might conduct key informant interviews with supervisors and their subordinates.

Other structure questions can be answered using **document reviews**. Tracking spreadsheets can be particularly helpful to organize information. The document or material is represented in a row (e.g., “staff training protocol”) and each column can represent the type of information you are collecting (e.g., “conforms to best practices,” “date of latest update/version”).

Process

Evaluating the processes of your program should include an analysis of administrative data from your organization (e.g., data on the number of individuals screened, the number of referrals made, and the number of service engagements). Before you can analyze these data, you first need a plan in place to collect and store the information. One way to collect participant data for screening and referral programs is to institute a practice of collecting basic demographic information from new clients via a short, standardized survey at every intake. There are many software packages for managing client information that vary in cost and utility, but the simplest solution is often a standard spreadsheet. Tracking spreadsheets, such as one created in Excel, will help you organize client information and easily tabulate and sort the data. Once a referral is made, that information should be similarly stored.

Outcome

Outcome data for screening and referral programs are usually quantitative and involve basic descriptive statistics, including frequencies, means, and standard deviations (e.g., patient demographics, referrals). Standardized screening tools should be used if you aim to assess changes in client symptomatology or recovery after using your program. For example, in a pre-post design for a peer-advocacy program, you might administer the Kessler 6 (K6) Psychological Distress Scale (Kessler et al., 2003) or a similar questionnaire. To show change over time, administer the questionnaire when clients first begin the program and then again

after they have been using the program for a set period (e.g., after attending a certain number of meetings with the peer advocate or on completing or exiting the program). For a quasi-experimental design, you would give the same questionnaire to clients who completed your program and to demographically similar individuals who did not go through a screening and referral program. Some common standardized assessment tools that may be useful for a screening and referral program are shown in Table 7.1.

Table 7.1
Example Standardized Screening Measures

Topic	Measure
Use of alcohol and other drugs	Alcohol Use Disorders Identification Test (AUDIT) (Saunders et al., 1993), Drug Abuse Screen Test (DAST-10) (Skinner, 1982)
General health and well-being	Columbia Impairment Scale (CIS) (Bird et al., 1993), Duke Health Profile (Parkerson, Broadhead, and Tse, 1990)
Psychological distress/depression	Kessler 6 (K6) (Kessler et al., 2003), Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer, and Williams, 2001), Geriatric Depression Scale (GDS) (Yesavage et al., 1983)
Trauma	Primary Care PTSD Screen (PC-PTSD) (Prins et al., 2003)

References

- Bird, Hector R., David Shaffer, Prudence Fisher, and Madelyn S. Gould, "The Columbia Impairment Scale (CIS): Pilot Findings on a Measure of Global Impairment for Children and Adolescents," *International Journal of Methods in Psychiatric Research*, Vol. 3, 1993, pp. 167–176.
- Kessler, Ronald C., Peggy R. Barker, Lisa J. Colpe, Joan F. Epstein, Joseph C. Gfroerer, Eva Hiripi, Mary J. Howes, Sharon-Lise T. Normand, Ronald W. Manderscheid, Ellen E. Walters, and Alan M. Zaslavsky, "Screening for Serious Mental Illness in the General Population," *Archives of General Psychiatry*, Vol. 60, No. 2, 2003, pp. 184–189.
- Kroenke, K., R. L. Spitzer, and J. B. W. Williams, "The PHQ-9: Validity of a Brief Depression Severity Measure," *Journal of General Internal Medicine*, Vol. 16, 2001, pp. 606–661.
- Parkerson, George R., Jr., W. E. Broadhead, and Chiu-Kit J. Tse, "The Duke Health Profile: A 17-Item Measure of Health and Dysfunction," *Medical Care*, Vol. 28, No. 11, 1990, pp. 1056–1072.
- Prins, A., P. Ouimette, R. Kimerling, R. P. Cameron, D. S. Hugelshofer, J. Shaw-Hegwer, A. Thrailkill, F. D. Gusman, and J. I. Sheikh, "The Primary Care PTSD Screen (PC-PTSD): Development and Operating Characteristics," *Primary Care Psychiatry*, Vol. 9, No. 1, 2003, pp. 9–14.
- Saunders, John B., Olaf G. Aasland, Thomas F. Babor, Juan R. de la Fuente, and Marcus Grant, "Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II," *Addiction*, Vol. 88, No. 6, 1993, pp. 791–804.
- Skinner, Harvey A., *Guide for Using the Drug Abuse Screening Test (DAST)*, Toronto: York University, 1982.
- Yesavage, Jerome A., T. L. Brink, Terence L. Rose, Owen Lum, Virginia Huang, Michael Adey, and Von Otto Leirer, "Development and Validation of a Geriatric Depression Screening Scale: A Preliminary Report," *Journal of Psychiatric Research*, Vol. 17, No. 1, 1983, pp. 37–49.

Counseling and Support

This chapter provides an overview of how to approach evaluation design and implementation for counseling and support programs, from designing a logic model, to identifying evaluation questions, to evaluation design and data collection.

Description of Activity

Counseling and support programs can take many different forms, depending on the target population and program goals. Examples of counseling and support programs include clinical counseling services provided at community wellness centers and family resource centers; parenting, home-visitation, and family-preservation programs; school-based bullying and violence-prevention programs; positive youth-development programs, such as after-school drop-in centers; and peer-to-peer support, senior-advocacy, and socialization programs.

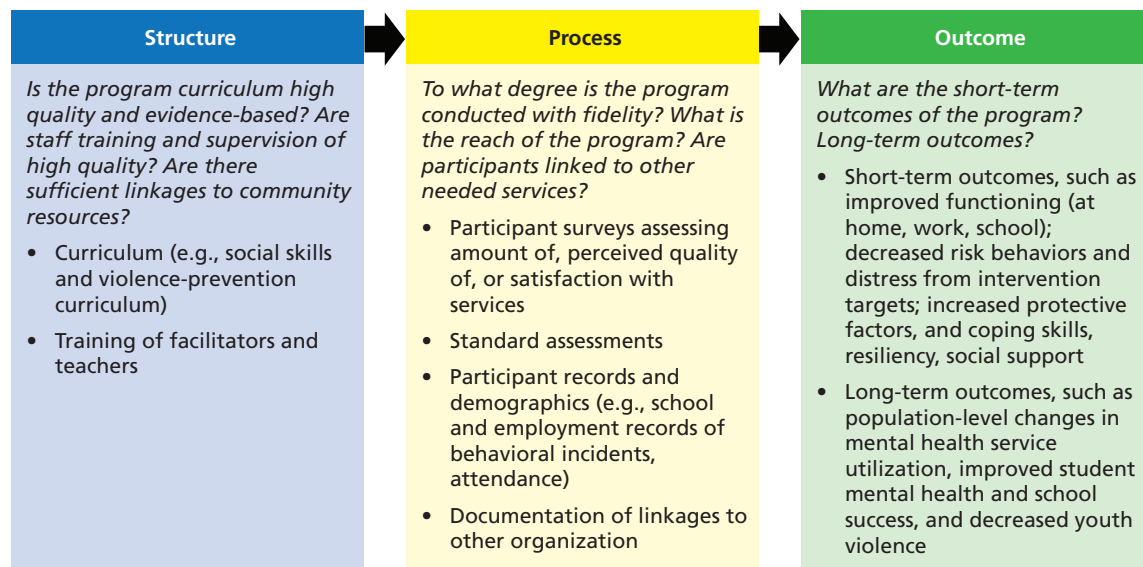
Logic Model

Figure 8.1 depicts a logic model for a counseling and support program. The structures relevant to a counseling and support program may include the program curriculum or the training provided to staff or supervisors. Processes are the quantity and quality of the counseling and support services, including the people reached. Short-term outcomes may include decreased risk behaviors; long-term outcomes are population-level changes, such as improved student mental health or decreased suicide rates.

Key Evaluation Questions

Below, we give several examples of structure, process, and outcome questions related to counseling and support programs. Your evaluation questions will depend on the type of information you are interested in and your organization's capacity to collect and analyze it. It is unlikely that an evaluation will be able to answer all of the questions suggested here. You will need to identify those that are most relevant to your programmatic needs. For counseling and support programs, evaluations are often centered on specific process questions (e.g., whether services are consistent with best practices) or outcome questions (e.g., the proportion of clients

Figure 8.1
Counseling and Support Logic Model



RAND RR1882-8.1

whose risk factors are reduced) that address how your program is supporting improved client well-being.

Structure

Some key questions to ask about the **structure** of your counseling and support program include

- **Is program curriculum of high quality and evidence-based?** Consider whether your program curriculum was developed in accordance with recognized best practices (such as those found in SAMHSA's national registry [SAMHSA, undated]), or whether best-practice standards are available for your population.
- **Are staff training and supervision of high quality?** Staff training and support are key elements that influence the quality of care that staff are able to provide. This question asks about the way your organization and program staff are structured.
- **Are there sufficient linkages to community resources?**

Process

Process evaluation questions look at the interactions between clients and counseling or support providers, including the quality of the services provided. Key **process** questions to ask about your counseling and support program include

- **To what degree is the program conducted with fidelity?** Having identified the target audience of your program, it is important to determine whether these people are actually the ones being reached. You will want to collect demographic data on clients who receive services through your program and compare this with program-reach goals.
- **Is the program functioning consistent with best practices?** This question asks whether the services provided are operating in a way that is consistent with best practices (such as

those identified in the section on structure evaluation, above). Note that, while best practices for screening and referral programs exist, they may differ (or might not have been established) depending on your target audience (such as sexual minorities, children, or veterans).

- **What is the reach of the program?** Knowing how many sessions are being delivered and how frequently the sessions occur can be useful to assess current demand or availability of counseling sessions, as well as for comparison with past numbers.
- **Are participants linked to other needed services?** When clients present with needs beyond mental health care (such as housing support or primary health care), are providers able to help clients access these services? Demonstrating how your program is or is not able to provide other supports may help you understand the full range of needs of your clients, as well as pinpoint how much of your staff's time and resources go to addressing these needs.

Outcome

Key **outcome** questions to think about include

- **What are the program's short-term outcomes?** A primary goal of counseling and support programs may be to improve functioning and to decrease clients' risk factors and mental health distress. Therefore, it is important to explicitly assess whether these outcomes are occurring. Another measure would be the proportion of clients increasing their protective factors (i.e., attributes that help mitigate the negative effects of adverse experiences), coping skills, and social supports. You may want to explicitly analyze individual protective factors, such as social support, to determine how your program is affecting them. Again, the factors you choose to assess may vary with the goals of your program (e.g., a socialization program might be assessed for client improvements in social support). Assessing changes in protective factors will help you build an evidence base for your program's positive effects or alert you to areas for improvement.
- **What are the long-term outcomes of the problem?** Long-term outcomes may include population-level changes in mental health service utilization, improved student mental health and school success, and decreased youth violence.

Evaluation Design

Structure

A **descriptive** evaluation design is often the most appropriate approach for answering questions about your counseling and support program's structure. The goal is to gain an accurate picture of the context in which counseling or support services are provided. Descriptive evaluation designs are ideal for assessing the structure of counseling and support programs because they are flexible and easily scale up or down, depending on your program size, needs, and evaluation resources. Because counseling and support programs tend to want to focus evaluation resources on process or outcome questions, a small-scale descriptive structure evaluation can provide a great deal of programmatic value for comparatively less effort. Descriptive methods can include a few **key informant interviews** with individuals who have direct experience in the program. These interviews can make it possible to obtain insider information about con-

nections with other organizations or the structure of staff supervision and support. **A review of organizational documents** can help you evaluate whether your program's curricula are in line with established best practices.

Process

Often, the best method for evaluating a counseling and support program's processes is also a **descriptive** evaluation design. Descriptive data can provide rich, detailed information to answer questions about the interactions between clinicians, providers, peers, or other program staff and the clients they serve. This design can answer questions about how clients access services for counseling and support, how treatment is delivered, whether the program adheres to best practices, and what resources clients are provided. Descriptive studies can scale up or down in terms of resource expenditures and complexity. Process evaluation questions about your counseling or support program may require more resources (time and funding) than the aforementioned descriptive structure evaluation; however, the value of the information is often similarly greater because of the opportunity to improve day-to-day interactions with your clients. A process evaluation for a counseling or support program will likely involve collecting data about clients and staff; generally, the more people involved in an evaluation, the more costly and time-consuming the process (hence the discrepancy between descriptive *structure* and *process* evaluations; structure evaluations for this kind of program tend to focus on tools, resources, and program staff, rather than service recipients). For counseling and support programs, a descriptive process evaluation is also an opportunity to view your program from the perspective of your client: When individuals walk through the doors at a community wellness center, what are their trajectories? Whom do they speak to? How welcome do they feel? What services are they offered? How culturally appropriate are those services? Because these are highly contextual questions, a descriptive design is often the best approach for capturing these interactions.

Semistructured interviews with staff and clients are one of the best ways to understand the interactions that occur within the context of your program activities. For example, interviews with staff or focus groups with clients can be used to evaluate whether the program activities are consistent with the best practices identified as part of the structure evaluation. **Reviewing organizational records** (e.g., attendance records, client data) can help you answer questions about who is being reached and how often program activities occur.

The data collection section in this chapter provides information on specific strategies to evaluate your program processes using a descriptive evaluation design.

Outcome

Evaluation designs that have a control group are the most robust method for assessing the effectiveness of an intervention. However, in community-based work, these designs can be impractical and prohibitively expensive. The next-best options are quasi-experimental designs and pre-post designs. For a counseling or support program, a quasi-experimental design could be achieved by comparing the outcomes of your program's clients with individuals who are not receiving counseling or support services. For example, if your program provides parenting support to at-risk parents and families, your control group could comprise demographically similar parents and families who are not enrolled in a parenting support program.

Pre-post designs are one of the more commonly implemented outcome evaluation designs by counseling and support programs. Many standard psychological assessments are used in

pre-post designs; for example, a community wellness center might administer the K6 assessment of psychological distress at an intake interview, and then again after two, three, or 12 months of program involvement (Kessler et al., 2003). While this pre-post design could not definitively state that the counseling or support received at the wellness center was responsible for changes in psychological distress, it would be a good indicator that a more robust evaluation design would be likely to show positive results.

Data Collection Plan

Structure

The data you collect will be used to answer the questions you identified in section about key evaluation questions. Typical data collection techniques for a descriptive evaluation of program structure are document review, observation, and key informant interviews. For example, you may be interested in evaluating the quality of the parenting curriculum that a home visitation program provides to new mothers. If your aim is to assess whether the curriculum meets current best practices, first research the guidelines you will be using for comparison. From these, create a checklist of important indicators. It can be helpful to enter them into a spreadsheet, with a column for each curriculum resource you are reviewing. Then, collect your program curricula and review them in relation to your checklist, making notes or simply check marks when guidelines are unmet, met, or exceeded. A document review could also provide information about how many trained peer specialists your program has and the quality of their training curricula.

The quality and consistency of staff training and supervision can be assessed through key informant interviews and observation. Interviews may be most useful if they are semi-structured—that is, consisting of a combination of closed-ended questions from which the interviewee picks a response and open-ended questions that encourage the interviewee to share opinions about a particular program aspect. For interviews, you will want to construct an interview protocol or survey that will be the same for each participant, allowing you to aggregate data across interviewees. Data can then be entered into a spreadsheet and easily tallied and analyzed.

Process

Many process questions about counseling and support programs are best answered by the people involved in the program, through key informant interviews or focus groups. For example, as part of your structure evaluation, you might have identified the best practices that your program intends to emulate; during your process evaluation, you may then be interested in the degree to which those best practices are adhered to in practice. Key informant interviews with program staff can provide firsthand detail about how counseling and support sessions are structured, as well as information about any challenges that staff may face in implementing best practices and work-arounds and strengths of the program.

For focus groups and interviews, you will want to construct an interview protocol. Semi-structured interview protocols contain open- and closed-ended questions and can provide you with detailed perspectives of program processes and quantitative information, respectively.

Other process data can come from client demographic data and attendance records. Creating an intake or initial demographic survey for new clients is a simple procedure that will

provide you with a wealth of process data. Data can be entered into a tracking spreadsheet (for example, with one row for each client). Ideally, attendance data for each individual will be entered into the same spreadsheet. Then you will be able to track who your program is reaching, identify demographic groups who may be underserved, and observe whenever there are patterns in who completes or drops out of the program.

Outcome

Outcome data are usually quantitative and reported in terms of descriptive statistics (frequencies, proportions, averages, and others). Standardized measures of client outcomes should be used to assess client outcomes for counseling and support programs, such as those shown in Table 8.1. For example, the K6 is a commonly used measure to assess psychological distress. In a pre-post design, clients would complete this measure when they begin the program and again when they conclude it. Comparing scores on standardized measures across demographic groups will allow you to see whether there are differences in outcomes by gender, race or ethnicity, language, or age.

School and employment records can also be used to track client outcomes (behavioral incidences and attendance). These may be more difficult to obtain at the individual level, and you may need to rely on self-reports. For example, an after-school support program for at-risk youth might track data on students' grade point averages; suspension and disciplinary referral rates; alcohol, tobacco, or drug violations; and involvement in extracurricular or college-readiness activities. Standardized measures could be used to assess whether youth identified or developed new strengths through their time in the program (assessed using the Devereux Student Strengths Assessment [LeBuffe, Shapiro, and Nagliari, 2009] or the Strengths and Difficulties Questionnaire [Goodman, 1997]); reduced drug or alcohol use and reported fewer impulsive reactions, depressive symptoms, and somatic complaints (e.g., using the Massachusetts Youth Screening Instrument–Version 2 [MAYSI-2] from Grisso and Barnum, 2001); or reduced suicidal ideation (Columbia Suicide Severity Rating Scale from Posner et al., 2008).

Table 8.1
Sample Outcome Measures

Topic	Measure
Protective factors (coping, resiliency, social support)	Devereux Adult Resiliency Scale (Mackrain, 2008), Devereux Student Strengths Assessment (DESSA-mini) (Nagliari, LeBuffe, and Shapiro, 2011), Family Quality of Life Survey (Hoffman et al., 2006), Lubben Social Network Scale (LSNS) (Lubben, 1998), Mental Health Continuum-Short Form (MHC-SF) (Keyes, 2002)
Psychological distress	Kessler 6 (K6) or Kessler 10 (Kessler et al., 2002; Kessler et al., 2003)
Comprehensive assessment (i.e., functioning, strengths, needs, risks)	Adult Needs and Strengths (ANS) (Praed Foundation, 1999), Child and Adolescent Needs and Strengths (CANS) (Lyons et al., 1999)
Behavior (youth)	Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997), Massachusetts Youth Screening Instrument–Version 2 (MAYSI-2) (Grisso and Barnum, 2001)
Suicide risk	Columbia Suicide Severity Rating Scale (C-SSRS) (Posner et al., 2008), Suicide Behaviors Questionnaire (SBQ-R) (Osman et al., 2001)
School climate	California Healthy Kids/School Climate Survey, DataQuest's Student Misconduct and Intervention data (California Department of Education, 2013)

Using a pre-post design, data would be collected at baseline—when the youth first joins the program—and at follow-up intervals. The frequency of follow-up measurements will depend on several factors specific to the program—youth attrition rates, ages or grade levels served, and program goals. If youth tend to drop out after six months in the program, follow-up assessments would need to be conducted at an interval that would catch the most youth but still have the possibility of showing results (in this example, at three or four months). If the program is limited to high school–age youth and most students join at the end of their junior year, the opportunity to follow up with youth may be limited to one year. Program goals may also dictate the follow-up or postassessment timeline (programs that have finite levels or steps that youth work through). In this example, you would want to pay particular attention to survey burden—surveys that are too frequent, too long, or too complex may result in inattention, frustration, and inaccurate results.

References

- California Department of Education, DataQuest, last modified November 8, 2013. As of April 28, 2017: <http://data1.cde.ca.gov/dataquest/>
- Goodman, R., “The Strengths and Difficulties Questionnaire: A Research Note,” *Journal of Child Psychology and Psychiatry*, Vol. 38, 1997, pp. 581–586.
- Grisso, T., and R. Barnum, *Massachusetts Youth Screening Instrument: Second Version (MAYSI-2)*, Worcester, Mass.: University of Massachusetts Medical School, 2001.
- Hoffman, L., J. Marquis, D. Poston, J. A. Summers, and A. Turnbull, “Assessing Family Outcomes: Psychometric Evaluation of the Beach Center Family Quality of Life Scale,” *Journal of Marriage and Family*, Vol. 68, No. 4, 2006, pp. 1069–1083.
- Kessler, Ronald C., Gavin Andrews, Lisa J. Colpe, Eva Hiripi, Daniel K. Mroczek, Sharon-Lise T. Normand, Ellen E. Walters, and Alan M. Zaslavsky, “Short Screening Scales to Monitor Population Prevalences and Trends in Non-Specific Psychological Distress,” *Psychological Medicine*, Vol. 32, No. 6, 2002, pp. 959–976.
- Kessler, Ronald C., Peggy R. Barker, Lisa J. Colpe, Joan F. Epstein, Joseph C. Gfroerer, Eva Hiripi, Mary J. Howes, Sharon-Lise T. Normand, Ronald W. Manderscheid, Ellen E. Walters, and Alan M. Zaslavsky, “Screening for Serious Mental Illness in the General Population,” *Archives of General Psychiatry*, Vol. 60, No. 2, 2003, pp. 184–189.
- Keyes, C. L. M., “The Mental Health Continuum: From Languishing to Flourishing in Life,” *Journal of Health and Social Behavior*, Vol. 43, No. 2, 2002, pp. 207–222.
- LeBuffe, P. A., V. B. Shapiro, and J. A. Naglieri, *The Devereux Student Strengths Assessment*, Lewisville, N.C.: Kaplan, 2009.
- Lubben, J., “Assessing Social Networks Among Elderly Populations,” *Family and Community Health: The Journal of Health Promotion and Maintenance*, Vol. 11, 1988, pp. 42–52.
- Lyons, J. S., G. Griffin, M. Fazio M., and M. B. Lyons, *Child and Adolescent Needs and Strengths: An Information Integration Tool for Children and Adolescents with Mental Health Challenges (CANS-MH) Manual*, Chicago: Praed Foundation, 1999.
- Mackrain, M., *Devereux Adult Resilience Survey*, Lewisville, N.C.: Kaplan, 2008.
- Naglieri, J. A., P. A. LeBuffe, and V. Shapiro, *Devereux Student Strengths Assessment—Mini*. Lewisville, N.C.: Kaplan, 2011.
- Osman, Augustine, Courtney L. Bagge, Peter M. Gutierrez, Lisa C. Konick, Beverly A. Kopper, and Francisco X. Barrios, “The Suicidal Behaviors Questionnaire-Revised (SBQ-R): Validation with Clinical and Nonclinical Samples,” *Assessment*, Vol. 8, No. 4, 2001, pp. 443–454.

Posner, K., D. Brent, C. Lucas, M. Gould, B. Stanley, G. Brown, P. Fisher, J. Zelazny, A. Burke, M. Oquendo, and J. Mann, *Columbia-Suicide Severity Rating Scale (C-SSRS)*, New York: Columbia University Medical Center, 2008.

Praed Foundation, *Adult Needs and Strengths Assessment: JCFS ANSA Manual; An Information Integration Tool for Adults with Behavioral Health Challenges*, version 2, Chicago, 1999.

SAMHSA—See Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration, National Registry of Evidence-Based Programs and Practices, undated. As of April 25, 2017:
<http://www.nrepp.samhsa.gov/landing.aspx>

Early Intervention Clinical Services

This chapter describes how to approach evaluation design and implementation for early intervention clinical services. After defining this activity, we review the major steps for this specific type of activity, including logic model development, identification of evaluation questions, evaluation design, and data collection.

Description of Activity

Early intervention services are targeted at identifying and treating early signs of mental or emotional health problems. These include early intervention clinical services for individuals at risk of depression, anxiety, trauma, or early onset psychosis (e.g., school-based trauma services, early identification and treatment programs for psychotic disorders), as well as programs that provide early childhood mental health consultations for children and parents of children with emotional or behavioral health problems (e.g., interventions for early childhood autism spectrum disorders).

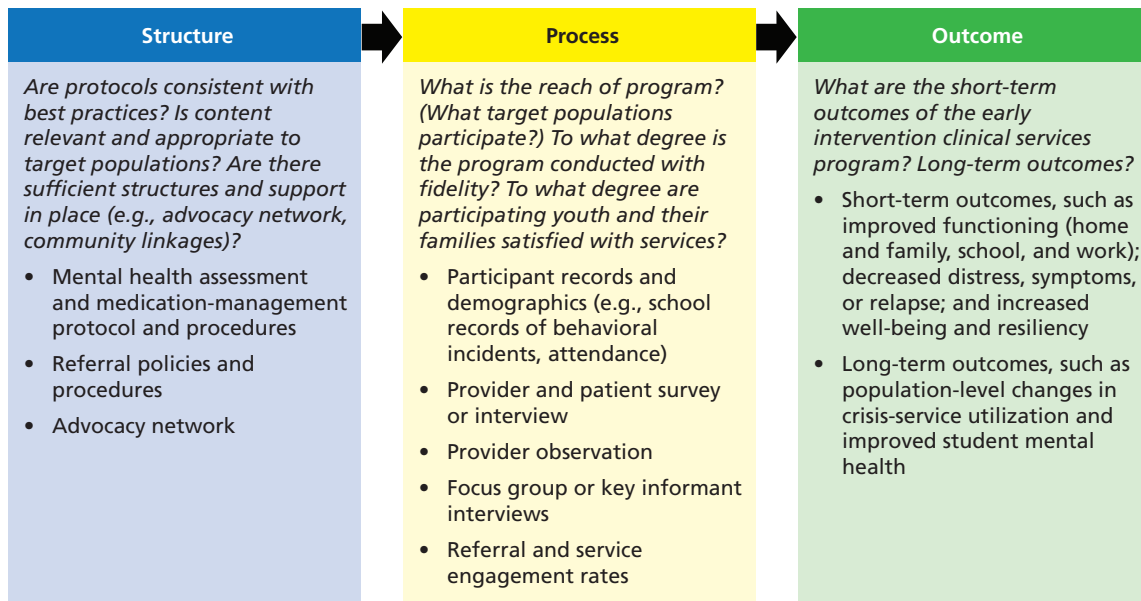
Logic Model

Figure 9.1 depicts a logic model to help guide the evaluation of your early intervention clinical service activities. The structure of an early intervention clinical service activity may include assessment protocols and procedures or the program's advocacy network. The process includes the demographics of service recipients and the fidelity to which providers adhered to clinical protocols. Short-term outcomes relevant to an early intervention clinical services or activity may be improved functioning at home or at school. Long-term outcomes may include population-level decreases in crisis-service utilization.

Key Evaluation Questions

Below, we give several examples of structure, process, and outcome questions for early intervention clinical programs. Your evaluation questions will depend on the type of information you are interested in and your organization's capacity to collect and analyze it. It is unlikely that an evaluation will be able to answer all of the questions suggested here, so you will need to identify those that are most relevant to your programmatic needs. For early intervention clinical

Figure 9.1
Early Intervention Clinical Services Logic Model



RAND RR1882-9.1

services, evaluations are often centered on specific process questions (e.g., whether services are available to traditionally underserved populations) or outcome questions (e.g., the reduction in the number of crisis events) that address how your program is supporting positive client trajectories through early intervention at critical times.

Structure

Structure evaluation questions refer to the context in which the early intervention clinical services are provided. Identifying the resources you have available will help you decide whether they are sufficient or whether there are gaps. Examples of important **structure** evaluation questions include

- **Are protocols and assessment tools consistent with best practices?** This question asks about the tools that clinicians use to make mental health assessments and manage medications, as well as organizational protocols and procedures for care. It is important to ask whether the tools that clinicians rely on are consistent with best practices to determine whether your program is meeting those standards.
- **Is the content of assessment tools relevant and appropriate to target populations?** Similar to the first question, consider assessing whether the content of the tools that clinicians use is appropriate for your target audience (e.g., if your program works with transition-age youth, are your assessment tools appropriately tailored to this age group?)
- **Are there sufficient structures and support in place (e.g., advocacy networks, community linkages)?** Examine whether the linkages that your organization has in place are sufficient for supporting your work.

Process

Process evaluation questions relate to the transactions between clients and providers, including the content, quality, and frequency. Key **process** questions to think about include

- **What is the reach the program?** This question asks about the demographic makeup of participants in your early intervention clinical services. To answer this question, you will first need to define what constitutes being *reached*. For example, are clients reached if they are aware of services, attend program activities, or benefit from services? Also, collecting demographic data about who accesses your early intervention services will give you an idea of who is using—and who may be underrepresented in—your program. You can then compare this information with your program’s target audience to determine whether you are serving who you intended to serve.
- **To what degree is the program being conducted with fidelity?** This question can be answered based on participant surveys and provider observations. It can also be considered relative to your organization’s operating capacity: Are there too few sessions because resources are limited? Are there more sessions than needed?
- **How satisfied are participants with the program?** What do participants say works for them? Is involvement in the services perceived as useful, beneficial, or worthwhile? Client satisfaction is important to measure because it often predicts future involvement and positive outcomes from the program and can be an indicator of areas for improvement (particularly if your program is experiencing low retention rates).

Outcome

Key **outcome** questions to think about include

- **What are the short-term outcomes of the early intervention clinical services?** A key outcome question to ask regarding your program’s effectiveness is whether clients experience reduced negative outcomes as a result of program engagement. For what proportion of participants has functioning improved and well-being and resilience increased? How many participants are seeing a decrease in distress, symptoms, or relapse?
- **What are the long-term outcomes?** Have there been population-level changes in crisis-service utilization? Has student mental health improved?

Evaluation Design

Structure

Typically, the most useful evaluation design for assessing early intervention clinical service structures is a **descriptive** one. A descriptive design allows you to take stock of the quantity and quality of resources available to your program, which might include clinical assessment tools and protocols or the program’s advocacy network. Descriptive evaluations typically involve **document reviews** and **key informant interviews**. A review of existing program documentation can be used to determine whether clinical protocols and assessment tools are consistent with best practices. Key informant interviews may be more useful to understand existing support structures and linkages to community organizations.

Process

Process evaluation questions for early intervention clinical services are often best answered using a **descriptive** evaluation design. A descriptive evaluation may consist of **document reviews**, **key informant interviews**, or **surveys**. A document review can be used to answer evaluation questions that draw on existing materials (for example, attendance or participation records). Key informant interviews or focus groups can provide rich, detailed information about program operations, client satisfaction, and potential areas for improvement. Surveys can be used to collect more-standardized information that can be tallied or analyzed across respondents.

Outcome

A **pre-post** design is often the most feasible evaluation approach for early intervention clinical services. In a pre-post design, outcomes of interest are measured when a client first begins the program and then again on completion of the program. The change over time can show improvements in client functioning. It is impossible to know for sure whether any positive outcomes are a result of the program because this design does not rule out other possible causes of improvement (e.g., because of typical emotional development). An **experimental** approach, in which a group of clients is randomly assigned to receive treatment and compared with a control group that does not receive treatment (or received treatment after the experiment ended), would allow you to make causal claims about your program (i.e., that the changes experienced by clients were a result of your intervention and not because of outside factors). This approach is more scientifically rigorous and may be necessary if your program does not already have an evidence base. However, it can also be more costly and time-consuming to implement. Where an evidence base already exists, we recommend implementing a pre-post design.

You will first need to determine what constitutes the *pre* and the *post* portions of your program. When does the program begin? When a potential client contacts (or is contacted by) your program or after the initial triage? While these two time points appear similar, they influence the data you will get and thus your interpretation of the program outcome. Individuals who contact the program may drop out before the first session. This could potentially highlight an area for program improvement (this would be more of a **process** evaluation question). You may have more opportunity to collect information from clients who attend the first session and thus be better able to compare them to those who complete the program. Some clients will likely drop out before completing the program. Consider collecting data at a midpoint (or midpoints) or collecting data at an exit interview. You will also need to define what constitutes program completion (the *post*).

Evaluation Data Collection Plan

Structure

A document review can help answer questions about whether your program's assessment tools and protocols are consistent with best practices and appropriate for your target audience. One way to approach this evaluation question would be to create a tracking document that lists the tools and protocols your program uses. Review the literature on existing evidence-based practices, or check a registry, such as the SAMHSA National Registry of Evidence-Based Practices

(SAMHSA, undated), and record whether each of your program's tools and protocols is present. Specifically, note whether evidence exists for the population of interest to your program.

A review of organization MOUs, contracts, and interagency correspondence can also provide information about linkages to other community organizations or advocacy networks. Key informant interviews and surveys with program staff are another way to determine whether your program's structures and supports are sufficient.

Process

Process evaluation questions can be answered using client surveys, key informant interviews, focus groups, or document reviews.

Surveys can provide standardized information that can be tallied or statistically analyzed. Focus groups and key informant interviews can provide rich, contextual detail about program processes. You will need to create a survey protocol for interviews and surveys. These can consist of open-ended or closed-ended questions. Generally, key informant interviews and focus groups will consist mostly of discussion points or open-ended questions, although you may want to ask participants a few closed-ended questions to assess demographics or general attitudes. Surveys usually consist of closed-ended questions, although you may also consider including short answer open-ended questions. When conducting interviews, be sure to consider power dynamics that may influence responses. For example, clients are often unlikely to report negative aspects of the program directly to program staff out of concern that it may affect the clients' services (a form of response bias). In this case, anonymous surveys—such as through a web-based survey program—may provide more-candid information. In-person interviews or focus groups can also be conducted by outside evaluators to minimize response bias.

Surveys can also answer questions about program reach. Creating an intake or initial demographic survey for new clients is a simple procedure that will provide you with a wealth of process data. Data can be entered into a tracking spreadsheet (for example, with one row for each client). Ideally, attendance data for each individual will be entered into the same spreadsheet. Then you will be able to track the people your program is reaching, identify demographic groups that may be underserved, and observe whether there are patterns in who completes or drops out of the program.

Questions about program availability (how often and where services were offered) can often easily be answered by collecting data from organization documents (e.g., event flyers, email reminders, training calendars).

Outcome

Your outcome data collection plan will depend on the evaluation design you selected. In the section on evaluation design, we recommended a pre-post design for early intervention clinical services, which involves collecting baseline and follow-up data on clients' psychosocial outcomes. Where possible, standardized assessment tools should be used to measure client outcomes (such as those provided in Table 9.1). Standardized assessments are those that have been scientifically tested to be valid and reliable measures of a particular construct (such as psychological distress or family quality of life). For example, for your mental health consultation program for youth with behavioral health concerns, you might consider using the MAYSI-2 (Grisso and Barnum, 2001) and the Devereux Student Strengths Assessment (DESSA-mini) (Nagliere, LeBuffe, and Shapiro, 2011) to assess pre- and postintervention levels of behavioral

Table 9.1
Sample Outcome Measures

Topic	Measure
Protective factors (coping, resiliency, social support)	Devereux Adult Resiliency Scale (Mackrain, 2008), Devereux Student Strengths Assessment (DESSA-mini) (Naglieri, LeBuffe, and Shapiro, 2011), Family Quality of Life Survey (Hoffman et al., 2006), Lubben Social Network Scale (LSNS) (Lubben, 1998), Mental Health Continuum-Short Form (MHC-SF) (Keyes, 2002)
Psychological distress	Kessler 6 (K6) or Kessler 10 (Kessler et al., 2002; Kessler et al., 2003)
Comprehensive assessment (i.e., functioning, strengths, needs, risks)	Adult Needs and Strengths (ANS) (Praed Foundation, 1999), Child and Adolescent Needs and Strengths (CANS) (Lyons et al., 1999)
Behavior (youth)	Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997), Massachusetts Youth Screening Instrument-Version 2 (MAYSI-2) (Grisso and Barnum, 2001), Youth Risk Behavior Survey (YRBS) (Centers for Disease Control and Prevention, 2016)
Suicide risk	Columbia Suicide Severity Rating Scale (C-SSRS) (Posner et al., 2008), Suicide Behaviors Questionnaire (SBQ-R) (Osman et al., 2001)
School climate	California Healthy Kids/School Climate Survey, DataQuest's Student Misconduct and Intervention data (California Department of Education, 2013)

concerns, as well as protective factors. Standardized measures can be used to answer outcome questions related to changes in symptoms and distress, well-being and resiliency, and behavioral functioning. The measures you choose should be administered in the same way at the baseline (pre) and follow-up (post) times. Data from these measures are usually reported quantitatively, in terms of frequencies, percentages, and means.

Other outcome data may be gleaned from school and employment records (for example, behavioral incidences and attendance). These data may be more difficult to obtain at the individual level, as you will need to carefully navigate Family Educational Rights and Privacy Act (FERPA) consent procedures or rely on client self-report.

References

- Brown, I., S. Neikrug, and R. I. Brown, *Family Quality of Life Survey*, International Family Quality of Life Project, 2000.
- California Department of Education, DataQuest, last modified November 8, 2013. As of April 28, 2017: <http://data1.cde.ca.gov/dataquest/>
- Centers for Disease Control and Prevention, "Questionnaires," data files, April 28, 2016. As of May 5, 2017: <https://www.cdc.gov/healthyyouth/data/yrbs/questionnaires.htm>
- Goodman, R., "The Strengths and Difficulties Questionnaire: A Research Note," *Journal of Child Psychology and Psychiatry*, Vol. 38, 1997, pp. 581–586.
- Grisso, T., and R. Barnum, *Massachusetts Youth Screening Instrument: Second Version (MAYSI-2)*, Worcester, Mass.: University of Massachusetts Medical School, 2001.
- Hoffman, L., J. Marquis, D. Poston, J. A. Summers, and A. Turnbull, "Assessing Family Outcomes: Psychometric Evaluation of the Beach Center Family Quality of Life Scale," *Journal of Marriage and Family*, Vol. 68, No. 4, 2006, pp. 1069–1083.

Kessler, Ronald C., Gavin Andrews, Lisa J. Colpe, Eva Hiripi, Daniel K. Mroczek, Sharon-Lise T. Normand, Ellen E. Walters, and Alan M. Zaslavsky, "Short Screening Scales to Monitor Population Prevalences and Trends in Non-Specific Psychological Distress," *Psychological Medicine*, Vol. 32, No. 6, 2002, pp. 959–976.

Kessler, Ronald C., Peggy R. Barker, Lisa J. Colpe, Joan F. Epstein, Joseph C. Gfroerer, Eva Hiripi, Mary J. Howes, Sharon-Lise T. Normand, Ronald W. Manderscheid, Ellen E. Walters, and Alan M. Zaslavsky, "Screening for Serious Mental Illness in the General Population," *Archives of General Psychiatry*, Vol. 60, No. 2, 2003, pp. 184–189.

Keyes, C. L. M., "The Mental Health Continuum: From Languishing to Flourishing in Life," *Journal of Health and Social Behavior*, Vol. 43, No. 2, 2002, pp. 207–222.

Lubben, J., "Assessing Social Networks Among Elderly Populations," *Family and Community Health: The Journal of Health Promotion and Maintenance*, Vol. 11, 1988, pp. 42–52.

Lubben, James, and Melanie Gironde, "Assessing Social Support Networks Among Older People in the United States," in Howard Litwin, ed., *The Social Networks of Older People: A Cross-National Analysis*, Santa Barbara, Calif.: ABC-CLIO, 1996, pp. 143–161.

Lyons, J. S., G. Griffin, M. Fazio M., and M. B. Lyons, *Child and Adolescent Needs and Strengths: An Information Integration Tool for Children and Adolescents with Mental Health Challenges (CANS-MH) Manual*, Chicago: Praed Foundation, 1999.

Mackrain, M., *Devereux Adult Resilience Survey*, Lewisville, N.C.: Kaplan, 2008.

Naglieri, J. A., P. A. LeBuffe, and V. Shapiro, *Devereux Student Strengths Assessment—Mini*. Lewisville, N.C.: Kaplan, 2011.

Osman, Augustine, Courtney L. Bagge, Peter M. Gutierrez, Lisa C. Konick, Beverly A. Kopper, and Francisco X. Barrios, "The Suicidal Behaviors Questionnaire-Revised (SBQ-R): Validation with Clinical and Nonclinical Samples," *Assessment*, Vol. 8, No. 4, 2001, pp. 443–454.

Petrillo, Giovanna, Vincenza Capone, Daniela Caso, and Corey L. M. Keyes, "The Mental Health Continuum—Short Form (MHC—SF) as a Measure of Well-Being in the Italian Context," *Social Indicators Research*, Vol. 121, No. 1, 2015, pp. 291–312.

Posner, K., D. Brent, C. Lucas, M. Gould, B. Stanley, G. Brown, P. Fisher, J. Zelazny, A. Burke, M. Oquendo, and J. Mann, *Columbia-Suicide Severity Rating Scale (C-SSRS)*, New York: Columbia University Medical Center, 2008.

Praed Foundation, *Adult Needs and Strengths Assessment: JCFS ANSA Manual; An Information Integration Tool for Adults with Behavioral Health Challenges*, version 2, Chicago, 1999.

SAMHSA—See Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration, National Registry of Evidence-Based Programs and Practices, undated. As of April 25, 2017:
<http://www.nrepp.samhsa.gov/landing.aspx>

Conclusion

There is growing understanding that, to improve the mental health of a population, programs should include a full spectrum of services, from prevention, to early intervention, to treatment. However, there is little information available about how to evaluate mental health PEI programs, as compared with treatment. At the same time, funders of mental health PEI programming are understandably demanding accountability. For example, in California, there is a mandate for counties to evaluate and report on their mental health PEI outcomes. Thus, it is critically important to use high-quality methods to evaluate mental health PEI programs to establish the value of continued investment in this kind of work. In addition, if standardized methodology and metrics are used across PEI programs being evaluated, it could be easier to compare outcomes for different programs, which facilitates decisionmaking regarding which programs to continue to fund. Furthermore, evaluation allows mental health PEI programs to see where they are succeeding and where there are challenges that need to be addressed related to program design, implementation, or reaching desired outcomes. Therefore, careful evaluation of mental health PEI programs facilitates quality improvement efforts.

We hope that this handbook is helpful in providing guidance to practitioners and organizations interested in evaluating mental health PEI programs and results in greater accountability, good decisionmaking regarding programming, and improvements in program quality.